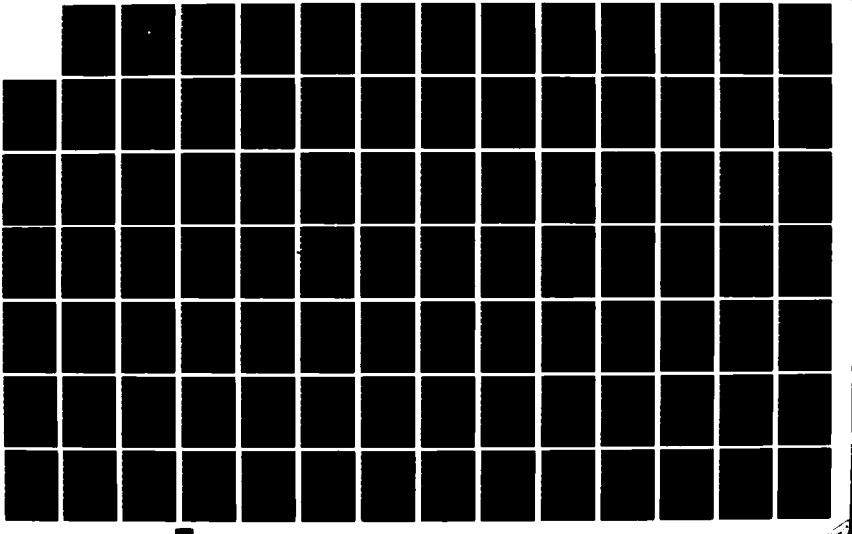
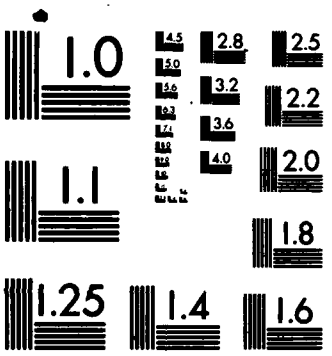
AD-A161 198 STRATEGIC PERFORMANCE MANAGEMENT EVALUATION FOR THE NAVY'S SPLICE LOCAL AREA NETWORKS(U) NAVAL POSTGRADUATE SCHOOL MONTEREY CA D D BLANKENSHIP APR 85 1/2

UNCLASSIFIED F/G 9/2 NL





NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

6 STRATEGIC PERFORMANCE MANAGEMENT EVALUATION FOR THE NAVY'S SPLICE LOCAL AREA NETWORKS

by

10 David D. Blankenship

September 1985

Thesis Advisor: Norman F. Schneidewind

Approved for public release; distribution is unlimited

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. A167198	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Strategic Performance Management Evaluation for the Navy's SPLICE Local Area Networks		5. TYPE OF REPORT & PERIOD COVERED Master's Thesis, September 1985 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) David D. Blankenship		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, CA 93943-5100		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, CA 93943-5100		12. REPORT DATE September 1985 13. NUMBER OF PAGES 138
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution is unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Network Performance Evaluation, LAN Performance, Network Performance Metrics, Stock Point Logistics Integrated Communication Environment (SPLICE), Supply Networks, Computer Performance Tools, Internetworking, Capacity Planning, Hyperchannel, Tandem Computers, Defense Data Network (DDN), Distributed Network Performance

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

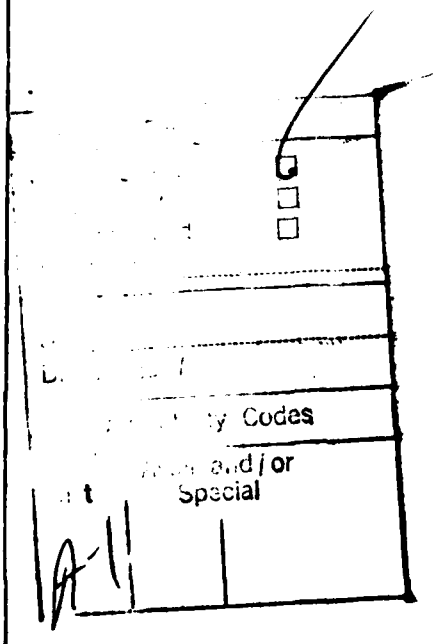
This thesis investigates those aspects of network performance evaluation thought to pertain specifically to strategic performance management evaluation of the Navy's Stock Point Logistics Integrated Communications Environment (SPLICE) local area networks at stock point and inventory control point sites. Background is provided concerning the SPLICE Project, strategic management, computer performance evaluation tools, computer and local area network performance metrics and performance evaluation (Continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

ABSTRACT (Continued)

methodology, capacity planning, the SPLICE LAN Communnications subnetwork hardware and software, and internetworking of SPLICE LAN's via the Defense Data Network (DDN). These topics, relevant case studies, and observations of one SPLICE LAN site are used to arrive at implications and recommendations applicable for improving future generic SPLICE LAN planning and performance.



Approved for public release; distribution is unlimited.

Strategic Performance Management Evaluation for the NAVY's SPLICE Local Area Networks

by

David D. Blankenship
Lieutenant Commander, U. S. Navy
B.A., Austin College, 1973
M.A., Austin College, 1974

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN INFORMATION SYSTEMS

from the

NAVAL POSTGRADUATE SCHOOL
September, 1985

Author: David D. Blankenship
David D. Blankenship

Approved by: Norman F. Schneidewind
Norman F. Schneidewind, Thesis Advisor

Barry A. Frew
Barry A. Frew, Second Reader

William R. Greer
William R. Greer, Chairman,
Department of Administrative Sciences

K.T. Marshall
Kneale T. Marshall,
Dean of Information and Policy Sciences

ABSTRACT

This thesis investigates those aspects of network performance evaluation thought to pertain specifically to strategic performance management evaluation of the Navy's Stock Point Logistics Integrated Communications Environment (SPLICE) local area networks at stock point and inventory control point sites. Background is provided concerning the SPLICE Project, strategic management, computer performance evaluation tools, computer and local area network performance metrics and performance evaluation methodology, capacity planning, the SPLICE LAN communications subnetwork hardware and software, and internetworking of SPLICE LAN's via the Defense Data Network (DDN). These topics, relevant case studies, and observations of one SPLICE LAN site are used to arrive at implications and recommendations applicable for improving the future generic SPLICE LAN planning and performance.

TABLE OF CONTENTS

LIST OF FIGURES

I. INTRODUCTION

A. THE PROBLEM

The contention that technology continually outdistances methodology is nowhere more accurately reflected than in the struggle by industry, government, and nonprofit institutions to get managerial control of local area network (LAN) technology. In LAN's a merging of hardware, software, and communications technologies has occurred and spawned new problems for the owning organization in how to optimize and provide for the evolution of a network so that benefits of hybrid technology can be reaped as a conscious effort rather than by accident. The issues at hand now include the merging and maturing of managerial skill along with our newly found technologies. During the same period when LAN technology was developing, strategic planning in management was being introduced formally as a way to not only provide sound decision-making for the current issues, but to anticipate any future environments and decision points as well. Management often has little time for formal planning.

The Navy's Stock Point Logistics Integrated Communications Environment (SPLICE) project for the Naval Supply Systems Command (NAVSUP) is an appropriate context to address these observations within the public sector. This phased plan to introduce policies and standards for future networking needs at multiple internetworked sites can continue to produce organizational benefits if some additional effort is expended now. This effort involves constantly assessing the past and present performance and anticipating the future workloads, technologies, constraints, and other factors in a total effort to provide

direction for the organization's network assets. At each SPLICE site a management team, consisting as determined by proper authority, of several appropriate members or as little as one dedicated person can play a crucial role in helping to avoid bad decisions, contribute to satisfied productive users at all levels, and get the most out of budgeted public funds. There is hardly a major corporation today without a performance evaluation division, or in more familiar terms, a capacity planning group. Bank of America represents a company with a transaction and online query environment as well as batch applications. In many ways that example is similar to the Navy's SPLICE system. This organization has a vice president in charge of capacity planning and separate divisions in charge of capacity planning for their TANDEM and their IBM systems. Personnel in these divisions are in addition to the programming and operations personnel. This illustrates how much value they place upon strategic performance evaluation in the form of capacity planning for information systems.

The Navy has, like industry, centralized this type of effort and made new application implementations, major hardware and software decisions, and performance studies from afar augmented by assistance visits to actual sites. Unfortunately, the Navy has many more SPLICE sites than most industries have computer installations, except perhaps for giants such as IBM. We do not fault this centralized approach. Rather, it is felt that a resident point of contact familiar with the particular hardware and software installation, the people, and the nature of that site's supply business can significantly assist in the success of centralized policy and standards and of the site's supply mission.

Each SPLICE node management team, because of the uniqueness of each node, must be able to not only measure

its own network performance, but to reach tuning, sizing, and capacity and configuration decisions for the future by interpreting those measurements. In practice, the Fleet Material Support Office (FMSO) conducts performance criteria and standards studies of each new application and an initial on-site performance evaluation at each SPLICE site. Local SPLICE node management should desire integral involvement in this process of establishing such a baseline of performance for their locally run applications. This performance evaluation experience gained while working with FMSO support groups can be valuable in assessing any future network or on-site modifications potentially affecting that particular SPLICE site.

In time as applications vary and as user demands accelerate, an organized ongoing methodology of strategically interpreting SPLICE monitored performance data will be essential in creating a historical data base, or at least a consistently documented approach to network performance management. Such a methodology can assist management of each SPLICE node in ensuring that SPLICE performance indeed matches the goals and objectives of the Navy's supply mission for SPLICE. It further seems logical that there should be some performance interpretation activities common to all SPLICE nodes and thereby applicable to any generic SPLICE node. This thesis addresses those potentially common computer network performance interpretation issues and suggests performance management guidelines which we believe to be relevant to the management of any SPLICE LAN node.

B. SPECIFIC THESIS OBJECTIVES

The primary objective of this thesis is to stimulate thought on how managers might usefully interpret local area

computer network performance data within the context of the organization's strategic plans and for the following purposes:

1. To improve network perfcrmance
2. To predict performance levels
3. To establish realistic performance standards and goals
4. To enhance network resources utilization
5. To assist capacity planning and configuration management decisions

Subsidiary research areas in support of this primary objective comprise the body of the remainder of this thesis. Additional issues to consider include the following:

1. The more deceptively simple decisions of which performance parameters to measure and interpret;
2. Current computer versus network qualitative and quantitative performance measurement concepts;
3. Consideration of ways in which internetworking SPLICE LAN nodes via the Defense Data Network (DDN) or other long-haul network will affect individual node performance.
4. Overlay of strategic management onto the LAN managerial environment.

Investigating these areas leads to questions of how to actually interpret network performance data assuming we know what data to gather and how to gather it. This question relates more directly to the primary thesis objective while remaining subsidiary areas support details of accomplishing this overall objective.

C. RESEARCH JUSTIFICATION

The case for strategic network performance management of SPLICE nodes is perhaps clearer than for such an activity in general. There exists a considerable body of professional literature on individual computer system monitoring. Evaluating the performance of an entire local area network of multiple processors, data paths, and connectivity through

telecommunications interfaces and protocols is much more difficult and less understood. Even less well-explored are the strategic management implications of such evaluated performance once it is obtained. Previous longterm guidance has primarily been accomplished on an ad hoc basis tailored to individual network situations. Despite the diversity of networks, there is a need for a generalized approach to strategically manage network performance so appropriate network resources are fully utilized and so management can retain a controlling as opposed to a reactive position. Beneficiaries of this research include not only the SPLICE operations and technical support managers, supply center ADP department heads, FMSO, and NAVSUP, but anyone desiring current research information on guidelines for performance management of LAN's.

D. SCOPE AND METHODOLOGY OF THE RESEARCH

This thesis has been narrowed in scope to necessarily strike a balance between conveying managerial guidelines and providing an essential technical foundation to the reader. This research is specifically limited in scope to applying concepts of strategic management and computer and known network performance evaluation techniques to operations of a generic SPLICE LAN. Various classes of network performance parameters will be discussed.

The following will <u>not</u> be covered:

1. Real-time operational network management
2. Performance tuning procedures or equipment
3. Casualty monitoring
4. Algorithms for processing or optimizing network routing directories
5. Excessive technical details of protocol considerations
6. Detailed software technical aspects beyond those needed in management implications of performance data

7. Details of the Navy supply system or its current policies or specific ADP transactions

8. sophisticated mathematical treatment of performance issues (queueing theory, modeling, etc.) User needs or procedures.

The emphasis is on long-term managerial interpretation of a variety of performance aspects in SPLICE.

The research involved a review of available NAVSUP/FMSO/Tandem Computer Corporation/Defense Communications Agency (DCA)/Federal Data Corporation (FDC) literature; an extensive survey of academic and professional book and article literature concerning performance of computer systems, networks systems, and network management issues; and on-site observation of a SPLICE LAN configuration at Naval Supply Center, Oakland, California with interviews of management responsible for implementing, operating, and evaluating SPLICE at that site. Information collected and conclusions drawn are primarily a result of exposure to primary and secondary source publications already mentioned, impressions from telephone or in-person interviews, and one on-site observation experience of two days in duration.

E. ASSUMPTIONS AND CAVEATS

The following series of assumptions and caveats have been made in producing this research effort:

1. It is assumed that tools and techniques of assessing individual performance for computer systems components can be applied to a degree to local area networks and their components. The element of synergy here will vary with the network and further research is needed.

2. References to quantitative specifics are for illustrative purposes only and make no attempt to imply a unique way of specifying user performance parameters. Any performance figures cited are likewise indicative of no particular SPLICE site or of any computer manufacturer. Such computations, graphs, or figures and accompanying discussions are to assist the reader in assimilating necessary facts to participate in a decision identification process.

3. All discussions of performance parameters, performance evaluation, and capacity planning will relate to the communications subnetwork elements of a SPLICE node (terminals, TANDEM FEP's, and HYPERchannel) and to the DDN influence on SPLICE performance. Specifically excluded are the SPLICE mainframes, mass storage devices, and the peripherals for FEP's. This is not to say that these components are unimportant to performance evaluation. Rather they will receive "black-box" treatment here. The FEP peripherals are simply considered a subset of the FEP in providing service.

4. The assumption is made that strategic management performance can be applied to various aspects of communications network performance evaluation.

5. Comments here address an installed, running system and not analysis or design issues.

6. This thesis aims at applying a narrow portion of network management, i.e. performance evaluation and planning, to SPLICE evolution in the future. The research results do not provide a cookbook of do's and don't's.

II. BACKGROUND

A. GENERAL OVERVIEW

The mission of the Naval Supply Systems Command (NAVSUP) is to provide effective logistics support to Naval fleet and shore commands [Ref. 1: p. 1]. NAVSUP formally initiated the Stock Point Logistics Integrated Communications Environment (SPLICE) through a tasking letter to Fleet Material Support Office (FMSO) on August 16, 1978 [Ref. 2: p. 1-1]. The project had been informally discussed since 1977. The Department of the Navy Code 041 (OPNAV-041) became the project sponsor. NAVSUP initiated SPLICE as a long-range four-phased project with the intent of augmenting the existing Navy Stock Point and Inventory Control Point (ICP) automatic data processing (ADP) facilities that support the Uniform Automated Data Processing System--Stock Points (UADPS--SP).

This augmentation was directed at the expanding problem of an unstandardized proliferation of unique hardware and software solutions to various new projects planned at numerous sites under (UADPS--SP), the projected ADP growth, and the need for state-of-the-art technical capabilities. Such unique solutions created the need for specialized hardware and software UADPS--SP interfaces from each new project multiplied in effect by the number of uniquely configured UADPS--SP sites. The UADPS-SP hardware, primarily the Burroughs Medium Size (B--3500/3700/4700/4800/4900) System at most sites, could not support multiple interfaces, projected increased service volumes, interactive processing requirements, and telecommunications functions simultaneously without a

significant redesign effort. There is an 8 to 10-year long-range plan to replace all ADP equipment according to NAVSUP [Ref. 1: p. 1]. This plan is the Stock Point ADP Replacement (SPAR) Project. SPLICE was to be one of the three shorter-term solutions using as much off-the-shelf capability as possible. SPLICE was to provide an effective and efficient standardized environment for absorbing communications workload from mainframe resources thus freeing them to handle increased applications volume, to support large scale interactive processing, and to serve networking requirements. Two other changes were to accompany this acquisition: replacement of older Burroughs mainframes with newer ones and replacement of outdated magnetic tape and disk drives [Ref. 1: p. 1].

SPLICE is now progressing with implementation, predominantly as individual unconnected nodes. The ultimate goal is to consolidate both local and long distance communications into a single integrated network using the DDN as a backbone [Ref. 1: p. 2]. The "foreground-background" processing concept of SPLICE is to be implemented at stock point sites using the Tandem Corporation minicomputer hardware and software suite with additional software supplied by FMSO. The initial field system prototype training and installation occurred at Navy Regional Data Automation Command (NARDAC) Jacksonville, Florida in July, 1984 thru January, 1985 [Ref. 1: p. 4]. A benchmark test is to be used according to plan [Ref. 1: p. 19] as the acceptance test for additional configurations which are ordered with sufficient components for the site workloads specified in the response document of the selected contractor [Ref. 3: pp. 9-1 to 9-198]. Local area network (LAN) performance requirements will likely skew from the original benchmark results because those results were based upon nonspecific pseudo-transactions specified in the

solicitation document [Ref. 4: pp. 70, 71] and which were likely to be processed at any SPLICE node. The dynamic character of new supply applications programs and the sheer volume increase in transactions, particularly interactive ones, will no doubt alter the current performance character and perhaps the desired performance requirements as well of each SPLICE LAN. Certainly, the performance of each SPLICE LAN will not match benchmark results exactly. Each SPLICE site, despite the adoption of standardized TANDEM equipment, will remain somewhat unique in terms of applications and transactions mixes and in some mainframe and peripheral hardware as well as in geographic dissimilarities. According to System Decision Paper III (SDPIII) [Ref. 1: p. 9], the ICP's presented a particular problem since they were IBM-supported and required TANDEM SNA software support in order to eventually be included in the SPLICE network. For these and similar reasons, all subsequent discussions will focus on considerations and actions from the viewpoint of management at any given SPLICE node.

B. SPLICE FUNCTIONAL AND CONTROL SUBSYSTEMS

The SPLICE concept was conceived to enhance the Navy's ability to continue both online interactive and batch supply order and communications processing through the advantages of internetworking LAN resources. The SPLICE functional requirements [Ref. 2] outlined the designs which were to be implemented by the system specifications [Ref. 5]. Because the SPLICE project has been ongoing since 1977 and is currently still in implementation stages, it is possible that numerous changes and modifications have transpired in hardware and software. The SDPIII [Ref. 1: p. 4] states that functional intent has remained fairly constant.

Each SPLICE complex will contain the same modular software subsystems. This reduces complexity, simplifies maintenance, and reduces the variety of interfaces [Ref. 2: p. 3-2]. The functional subsystems of the foreground are discussed at length in the SPLICE functional description [Ref. 2: pp. 3-3 to 3-11]. These functions are as follows:

1. Terminal Management Subsystem -- three components which provide the terminal handling, security, and user process selection
2. Transaction Support Processing Subsystem -- eight components which provide user entry points into the various transaction processing services of SPLICE
3. Complex Local Computer Network (LCN) Control Subsystem -- provides the physical and logical connection to the LCN
4. Site Management Subsystem -- three components which provide access to the system for the System Administrator, the console operator, and the CRT user
5. Internal Management Subsystem -- four components which control internal routing of all data and files destined for LCN/Data Communications Network (DCN)/terminals, interpretation and execution of command messages, and system monitoring
6. Data Exchange Subsystem -- three components which control data set files entering and leaving the site, queue files of backlogged transactions, and site peripherals
7. Site DCN Control -- two components which support the communications interface, control, priority, workload leveling and logging of output traffic.

The same basic functions somewhat distilled are presented in Federal Data Corporation's (FDC) contract award in slightly different names with overlap existing so that it is not possible to make a one-to-one correspondence. As cited in a more recent contract award through FDC [Ref. 3: pp. 10, 11, 13] these functions are as follows:

1. Terminal Management,
2. Batch Processing,
3. Data Set Management,
4. Peripheral Management,
5. Complex Management, and
6. SPLICE FEP support.

C. SPLICE LAN ARCHITECTURE

The configuration architecture for a representative SPLICE LAN will now be briefly presented. The node referred to directly or by implication here will resemble Naval Supply Center Oakland, California more than any other since that node was visited during the research phase of this thesis to gain on-site exposure to the site configuration and environment. The many functions, subsystems, and vendor equipment capabilities have been explored and reported in other works, including NAVSUP's own functional and system specifications documents and research work conducted by several faculty and graduates during the last three years at the Naval Postgraduate School in Monterey, California. The reader is directed to these works for detail beyond the scope of this research. Only a brief description of the LAN configuration will be covered here to set the stage for later discussions of the SPLICE communications subnetwork and its performance.

Basically, the stock point nodes can be described as a flow from the user through the communications subnet to the node mainframe(s), to the FEP itself, or to internetworked sites. The online terminals are connected in groups of six to a common modem which connects to a coaxial cable. The cable runs to a TANDEM frontend processor which routes traffic either locally as "pass-through" to the Burroughs (or IBM mainframe at ICP's) via a HYPERchannel high speed local network, processes the traffic as necessary at the TANDEM processor cluster, or routes the traffic in gateway fashion to the DDN. At the ICP's, of course, the terminals and mainframes may differ; however, the TANDEM FEP will remain a standard for all SPLICE sites.

D. STRATEGIC PLANNING

1. The Strategic Planning Discipline

The views of distinguished writers in the field of strategic thought best convey a feeling for strategic thought and process. These will be generously used here to reduce the amount of material which would otherwise have to be explained as a background for performance evaluation. In the case of SPLICE or any other LAN management, a plan is essential simply because of the investment at stake and because managers can no longer make their way without some external knowledge of the environment affecting their decisions. It is noteworthy that NAVSUP has, after the SPLICE project inception, approved a Strategic Planning Document [Ref. 6] for the SPLICE organizational strategic plan. Radford aptly put it this way:

> "Despite intuitive capabilities of successful managers, the increasing complexity of their environments places increasing demands upon them. It is more difficult to encure all necessary factors are included in a strategic plan unless a basic structure is adhered to beyond mere intuition." [Ref. 7: p. ix]

According to Radford, the aim of strategic management is as follows:

> ". . . . to ensure present and future activities of an organization are appropriately matched to environmental conditions under which the organization operates . . . to select future activity and action courses for the organization which will result in a high degree of achievement of objectives." [Ref. 7: p.4]

The process described by Henry Fayol is as follows:

> ". . . . (a) visualizing possible future situations in which the organization concerned might be involved, (b) placing these situations in an order of preference relative to the objectives of the organization, and (c) considering ways in which the most preferred of the future situations considered can be brought about and the least preferred avoided." [Ref. 7: p. 1]

Consider 62 or more separate interconnected SPLICE LAN's in separate geographic areas with increasing volume usage of increasing numbers of application processed not only locally, but upon demand at other nodes as well. Add to that a multi-vendored technology which cannot handle further expansion and a few irregular budgetary constraints or regulatory constraints and you have a hostile environment.

2. Strategic Planning Characteristics

To implement strategic planning within an organization one must recognize what constitutes strategic planning, what it can be applied to, and its limitations. Since the external environment affects the entire organization, it most probably touches all activities of the organization. Performance evaluation and interpreting that evaluation for capacity planning are activities needed in a LAN organization. Radford writes:

> ". . . . strategic planning provides . . . a set of strategies and policies that constitute a framework for planning and decision-making throughout the organization . . . . They are . . . extensions and amplifications of the organizational objectives on which the . . . (planning) . . . process is based . . . . This planning process must keep in mind (1) the mission of the organization, (2) the objectives of the organization, and (3) values and preferences of the organization . . . ." [Ref. 7: p. 4]

This indicates that strategic plans have a way of communicating organizational objectives to decision makers. This is a desirable way to communicate performance goals and standards throughout a SPLICE site. It might be necessary to remind the reader that an organization's strategic plan may be dictated from higher authority levels, but performance evaluation can still be a relevant part, or even added at the local level.

Strategic planning is highly subjective and unlike controlled environments, results of strategic planning cannot be compared with what might have transpired without it. It is not only a long-range view. Many times short-term factors arise as a result of unpredictable external events. This causes a need for change or modification of future directions and activities. Its application can be broad and applied to almost any unstructured situation. It elicits consideration of alternatives, stimulates discussion and communication, creates a framework for decision-making, and nourishes the mechanism for responding to change. One limitation is that it provides a range of possible reactions to future conditions and not "the answer". Another limitation is that strategic planning is iterative and must be continuously reviewed (not only at fixed intervals). Strategic planning is a procedure for recognizing risk and taking advantage of it, not eliminating risk altogether. Strategic decisions are often unique and not amenable to analytical formulations, such as in structured situations. Hence, modeling and simulation can play key roles. [Ref. 7: pp. 4-7, 9]

Radford [Ref. 7: pp. 12-13] offers the following four components for describing the procedure of strategic planning:

1. "Review mission and objectives.
2. Consider existing and future decision situations.
3. Plan for implementation.
4. Review and reappraisal."

Harry Katzan applies strategic planning to local area networks by advocating a three-point strategy of assessing the current position, (Where are we?), setting goals (Where are we going?), and direction (How do we get there from here?). Direction is emphasized as the major

component. He views LAN's as potentially unstructured operating environments requiring a high degree of integrated planning in application functions, media, "products" (network components, peripherals), and vendors. [Ref. 8: pp. 164-166]

An interesting closing note on strategic planning characteristics is that the period during which collapse or disaster develops is of the same order as the time span into the future with which such planning studies are concerned. Not all calamities develop so gradually, but even in technologies such as LAN's there is adequate preparatory time. A key point to focus upon is that here we want to apply strategic planning principles to a narrow aspect of network management, i.e. performance evaluation, and to keep in mind that this includes far more than capacity planning alone.

B. STRATEGIC PLANNING FOR SPLICE

The objective of this research is to apply the strategic planning discipline to the results of measuring and predicting network performance so SPLICE management can correctly interpret current network activity and prepare for future demands. In SDPIII interpretation is described as having been approached in a somewhat foreseeable preplanned manner. Following the initial installations, a series of upgrades at each site have been planned according to projected site application implementations and workload growth. The contractor, under the indefinite delivery and quantity contract terms, is encouraged to suggest improvements and substitutions which might ennance performance. These are separate from scheduled upgrades. The ccntractor is only bound to provide modular architectural units according to the initial benchmarked

configuration sizing requirements for handling the current and near-term projected workloads at each site. The subsequent upgrades of additional units are scheduled in the contract to ensure fixed prices and contractor commitment. If the upgrades are insufficient to handle the proposed worklcads mentioned in the contract, then the contractor furnishes additional equipment at no cost. If the workload exceeds that proposed, then negotiations of equipment amounts and costs are undertaken. The question at issue here is that while this basic approach ensures SPLICE has some contractual flexibility and planning for capacity needs, the fine-tuning of an ongoing performance evaluation activity by each unique SPLICE site management is largely avoided. Without a concurrent effort to evaluate actual performance over time at each site, some sites may end up with delayed application implementation and excess capacity for a time which costs the government. At other sites, unforeseen workload may force the government into expensive additional contract negotiations. The conclusion offered is that while SPLICE is apparently well-prepared in terms of Radford's first three procedures of strategic planning, the fourth procedure could be better carried out through a continuing performance evaluation effort which reflects organizational objectives. This research, or modifications of it, could easily be integrated into a portion of NAVSUP's existing SPLICE Stratgic Planning Document. By any other name strategic performance evaluation management would still be recognized as perhaps what industry has referred to as capacity planning for some time now. We return to it again in chapter V after exploring in chapter's III and IV how and what quantities to measure in order to get a perception of performance in computers and networks.

III. CONCEPTS IN COMPUTER AND NETWORK PERFORMANCE

This chapter serves as a foundation and a technical terminology bridge in moving from a general discussion of SPLICE and strategic management principles to the more particular goal of strategically evaluating or interpreting performance of SPLICE. A detailed glossary is found in Appendix A to provide the reader with necessary technical detail and to facilitate explanation.

A. DEFINITION AND PERSPECTIVES ON PERFORMANCE AND PERFORMANCE EVALUATION

Before perfcrmance can be evaluated, its nature must be defined. "Performance" could be described as the observed behavior (in discrete units or in general) of a system in a certain situation as compared to some predefined criteria or measurement. Ferrari supports this definition [Ref. 9: p. 10]. He likens "observed behavior" to measured characteristics of the physical system, "a certain situation" to operating conditions of the system at evaluation time, and "predefined criteria or measurement" as performance indices. Measurement is a key element of determining performance. Measurement is basically collecting information about some system as it is used or as it operates. We measure to determine performance. The "system" explored in this thesis is a hybrid network incorporating two of the three types of local networks, the local area network (LAN) and the high speed local network (HSLN). These two types co-located in nodes are connected to similar hybrid nodes via some long-haul network (LHN) such as the DDN.

When networks are involved, the definition of performance must become more specific with respect to exactly what behavior and what physical aspects are measured in a network sense. In network performance, we are no longer talking only about discrete processors, disk units, or programs. In a network we see a distributed entity made up of many components connected together through communications links for the purposes of resource sharing, exchanging message and data traffic, reducing the effects of distance, and providing a variety of services to users. In network performance there are additional entities which enter into an assessment of network performance, such as protocols, telecommunications connections, frontend and backend processors, high speed bus or ring connections, and nearly always, more remote terminals than in a multi-terminal mainframe situation. Network performance is not only highly dependent upon all these elements, but upon their mutual interactions as well. Ferrari [Ref. 9: p. 1] indicates that performance refers to how well the system provides all designed facilities to a user. Unfortunately, the definition does not get any better, and one must realize that a notion of performance is heavily dependent upon the context. Its factors are in large measure qualitative rather than quantitative. Borovits and Neumann [Ref. 10: p. 3] contend that performance has no meaning unless it refers to a specific application. If that were strictly true, how does one speak of performance of a network (or of a computer for that matter) where many applications may be in process concurrently? Performance results depend upon the interaction of many things, including software, transaction or application mix, amount of monitoring requiring system assets, quantity of users during a time period, and overhead for reasons other than monitoring, as well as system configuration.

A better understanding of the intent of this thesis and its perspective on performance evaluation can be obtained by noting its relationship to Stallings' framework of network management and to performance evaluation objectives or purposes set forth by both Ferarri and by Borovits and Neumann [Refs. 9,10: pp. 2, 6-7]. Stallings' definition and framework for network management appears to be the most detailed and comprehensive. Other authors [Refs. 11,12: p. 86, 54] tend to restrict the definition to primarily real-time operational concerns such as monitoring, fault management, configuration management, load balancing actions, and reporting. Most authors offer very little about wider reaching aspects involving such concerns as planning, security, data bases, and performance interpretation of data gathered about network activity.

Most publications on network subjects lean heavily toward design of network topology, issues of optimum design for the user's needs, protocol issues, and monitoring to improve current performance. Studies in the area of capacity planning have been one exception to this short-term view. Stallings' definition of network performance is as follows:

> "Network management is a broad concept that encompasses those tasks, human and automated, that "support" the creation, operation, and evolution of a network . . . it is the "glue" or infrastructure of techniques and procedures that assure the proper operation of a system." [Ref. 13: p. 326]

He qualifies this definition by indicating that "support" should not be interpreted to mean the functions or disciplines involved in controlling development or ongoing use of a system. However, the words "evolution" and "proper operation" in the definition certainly seem to imply some sort of performance assessment and managerial intervention

to ensure the evolution is controlled and in concert with organizational objectives. For this reason, network management can be viewed in a much broader scope. This broader scope will be pursued here.

There is an inconsistency in trying to restrict and narrow a definition of network management on one hand and viewing it as a "broad concept" as Stallings has on the other hand [Ref. 13: p. 328]. This inconsistency can be seen in the functions Stallings ascribes to network management:

1. Operations --day-to-day operational status of the network, including traffic and performance status, active devices and accounting and billing.
2. Administration --managing the use of the network through system generation, passwords control, resource and file access management, and administering an appropriate charge-out system. (It can be argued that matters of budgets, personnel and staffing, auditing, accounting, and training are general management features. But for inclusiveness, they are included here with administration.)
3. Maintenance --detection and reporting of problems through human or automated means to assure that the network continues to operate.
4. Configuration Management --management of the system's hardware and software life cycles and its evolving configuration by tracking, documenting and controlling changes to, maintaining status on, and ensuring the continuing adherence to requirements by all components.
5. Documentation/training function --educational functions for developing and maintaining documentation.
6. Data base management --provide updating and care management of the network management data base.
7. Planning --providing for ongoing requirements analysis, configuration change, and capacity planning.
8. Security --protect against prevention and detection of unauthorized network access.

Clearly some of these functions omit the disciplines involved in developing and modifying a system, but do not omit whatever managerial functions are involved in controlling network development and evolution. The problem may be more that semantics. It is difficult to say when

Stallings' "network management" ends and strategic interpretation of phases and outputs of that management process begins. While the first three functions above comprise the responsibilities of the Network Control Center (NCC), we believe, the NCC's role, like the definition of network management, can be extended. The NCC's role should include aspects of security and even portions of configuration management in a short-term sense. Stallings elaborates on the functions of the NCC which he depicts as primarily operational or maintenance in nature: configuration functions, monitoring functions, and fault isolation. The monitoring functions of an NCC can be further decomposed into performance measurement (gathering data), performance analysis (data reduction and presentation), and synthetic traffic generation (observing the network under a hypothetical load). In these activities lie some sources of the performance data we seek to evaluate throughout the network's life cycle.

The four performance evaluation objectives Ferrari describes are very similar to those of Borovits and Neumann. Each author implies that performance evaluation of a system is necessary throughout the life cycle, and not merely after it is installed. The perpetual objectives outlined by all three author groups above and characterized in Ferrari's terms [Ref. 9: pp. 2, 3] are these:

1. Procurement --This includes all evaluation problems associated with a choice of a system or components among alternatives which matches the conceived workload.
2. Improvement --This includes any performance evaluation problems which occur in existing operational systems.
3. Capacity planning --This objective refers to the prediction of when the current system capacity will become insufficient to process the required workload at a given level of performance and thus require modular or complete replacement.
4. Design --This includes any performance problems associated with designing an appropriate system.

The names of the phases vary, but the essence of the cycle is captured by all the authors. The four areas of evaluation are referred to by Morris and Roth [Ref. 14: p. 10] as phases. They are named Procurement, Installation, Operation, and Transition. While Morris and Roth's "Procurement" and "Transition" are easily identified as Ferrari's "Procurement" and "Design", respectively, Ferrari has no parallel for Morris and Roth's "Installation". An argument can be made that this represents a genuine phase although a relatively short one. Morris and Roth then lump Ferrari's "Improvement" and "Capacity Planning" into the single "Operations" phase.

Of particular relevance to this thesis is the application of performance evaluation to the SPLICE context within a broad definition of network management. Emphasis is upon the improvement and capacity planning objectives stemming from interpretations of those three components of performance monitoring cited above. The monitoring functions will be assumed here to be complete and available. It is the interpretation of reduced data from monitoring and from the results of application of performance tools and techniques which we will concern ourselves with here.

B. WHY PERFORMANCE?

As stated by Abrams [Ref. 15: p. 313], most research study has focused upon the individual performance of components of computer and communications networks such as computers themselves, disk drives, high speed data channels, software programs, network switches, and so on rather than functioning networks in toto. Three salient trends have caused a surge in the need for accurate, even reliably approximate, methods of estimating the overall performance of a network. The Auerbach Management Series [Ref. 16: p. 1] mentions the following trends:

1. the rapidly maturing network technologies,

2. the demands that upper level management and the users are placing upon information systems management to achieve some means of evaluating and predicting network performance, and

3. the recognition that there is an important link between user productivity and system performance. There should be no obstacle to building an understanding of network performance based upon previous studies of discrete components. In fact, this move from a micro to a more dynamic and synergistic macro level can be undertaken with a little less apprehension since modification to or direct use of tools and methods used in component studies may hold promise for network use.

Another reason for the desire to assess performance of systems is cost. Even in nonprofit or government situations where costs may be perceived as secondary to mission, that concern of management seeking the best cost-performance ratio possible is still present. Any information systems manager, even if not concerned in the near term with possible replacement of equipment, will nevertheless, seek to get as much benefit as possible out of currently installed hardware and software. In the context of SPLICE, one of the main purposes of measurement is to aid in the evaluation of service provided to the terminal user. Here the link between system performance and user productivity becomes evident.

C. WHAT SHOULD BE MEASURED?

Before any system's performance can be correctly evaluated, there must be some agreement upon what entity we are attempting to take measurements upon and what aspects of the entity are necessary to measure and interpret. In the absence of agreement about the specific measures to make, at least we need some rational approach to justify what we measure or to explain deviation from any accepted set of standards. In computer performance, such things as paging rates, throughput, input/output channel usage, turnaround

time for batch mode and response time for online mode, percent of cpu utilization, component utilization, availability and reliability are of interest. Some or none of these may be pertinent to a particular network. There is currently no standardized set of performance metrics for networks. For that matter, there are no universal standards for computer performance standards either. Ferrari [Ref. 9: pp. 11-33] lists and describes some more commonly accepted ones.

Computer or network performance personnel attempting performance comparisons between LAN architectures or seeking to develop a performance evaluation program are often frustrated and certainly hampered by lack of standardized metrics. Assuming that standardized metrics are necessary, there are at least three problems which emerge immediately according to Amer and Goel [Ref. 17: pp. 195-196]. These are the following:

1. Performance metrics are not always defined in a precise, unambiguous way. In fact, they have generally been inconsistently defined, thus preventing users from specifying their requirements precisely and unambiguously.

2. There is often no distinction between user-oriented and network -oriented paramenters.

3. While it is acknowledged that some parameters will have meaning only for certain technologies, topologies, or protocols, every effort should be made to make performance metrics as independent of such associations as possible. Without such independently defined metrics, efforts to correlate studies performed on LAN's or to compare the performance of different LAN topologies will continue to be difficult.

Four metric attributes allegedly desirable are claimed by Amer and Goel to apply to topology independent metrics; however, the simplicity of these attributes does not prevent them from applying to particular topologies and protocols such as by rollcall or CSMA networks. The four attributes these two authors describe [Ref. 17: p. 196] are as follows:

1. User orientation --metrics should describe performance characteristics relevant to the network needs of users without measuring user perfcrmance.

User orientation refers to the metric which describes performance of service to the end user while user effect describes the effect of user interaction with a specific network.

2. Simplicity --in order for users to precisely and unambiguously define their performance requirements they must be able to clearly understand and communicate definitions of metrics.

3. Minimal overhead --metrics should be measureable without imposing an excessive amount of overhead on the system. For this reason, bit-level metrics may be impractical, especially in a LAN environment, and of little benefit to a user.

4. Comprehensive —metrics should encompass all aspects of performance significant to data communications users.

As previously alluded to, even when a standardized set of metrics has been accepted there will be other measurable aspects particular only to one topology, for instance, which users or management will desire to quantify and study. Therefore, the four attributes above must be balanced by two additional factors mentioned by Ferrari [Ref. 9: pp. 9-10] as follows: (1) The projected type of information required by management may dictate which performance measurement parameters are necessary. This may be required in spite of the best intentions to standardize metrics. (2) The type of network technology being monitored may bias or otherwise restirct accurate performance metrics or may even make it useless to gather data on one metric for a particular LAN technology which is vital to another.

A few pertinent examples serve to illustrate these two factors. An interoffice automation and broad service spectrum LAN in a major corporation might be valued for its availability and reliability from a user standpoint, while an interstate bank supporting hundreds of automatic teller machines (ATM's) in LAN's may be more concerned with interactive terminal productivity and response time for customers. The management of SPLICE LAN's may be much more concerned with interactive terminal productivity and require that throughput and minimum network delay for the user be

the primary parameters of concern. Transfer rate is another metric which might be of concern for file transfer and batch mode operations in a communications net.

More on specific network metrics and SPLICE network evaluation will be covered in Chapters IV and V, respectively. For now the subject of network performance metrics has only been introduced.

Published articles in recent years have referred to loosely defined parameters of indices coming into some general acceptance as worthy of measuring. The ultimate decision will, of course, reside with the organization base upon its perceived needs. Whatever the position with regard to any universally accepted standard, the important issue is that an organization adopt some standards as a basis for trend analysis and to relieve confusion. As one might guess, new technologies coming into the network arena and the merging of technologies such as will be seen in the integrated services data network (ISDN) concept where voice, digital, and video data may be transported over a common medium will complicate the case for metrics and perhaps make it even a more vital issue.

D. HOW DO WE MEASURE OR EVALUATE PERFORMANCE?

1. Computer Performance Evaluation Tools in General

Many of the tools and lessons of computer performance evaluation (CPE) should not be ignored in attempting to estimate network performance evaluation (NPE). Ideally, important metrics should calculable from existing hardware and software data collection systems/tools/techniques already available to a particular site. This may not be practical and special tools may be necessary, even vendor's. Whether calculations based on data gathering alone are sufficient for properly assessing

network performance behavior is an issue best debated elsewhere. For now, as a matter of background we are concerned with CPE. Morris and Roth [Ref. 14: p. 2] see CPE as the application of special tools, techniques, and analytical methods to improve the efficiency or productivity of existing or planned computer installations. Where and how we measure are largely dependent upon which type of tool is used. There are nine generally recognized CPE tools/techniques which will also be considered for use in some way for network performance evaluation, either singly or in combination. Morris and Roth [Ref. 14: p. 6] view CPE tools as fitting into two categories:

1. measurement or
2. predictive

A brief description of each tool or technique and advantages and disadvantages can be found in Appendix B.

While it may be restated later, the importance of not relying on any single tool universally cannot be overstated. Morris and Roth's [Ref. 14: p. 10] life cycle phases for systems and the tools appropriate for tasks in each phase bear this out. No one tool is a panacea, nor can any tool be applied at random to every situation. The tools employed must fit the case. Some suggestions for which tools might be appropriate for SPLICE appear in Chapter IV.

The reader should realize that in heterogeneous LAN's such as SPLICE where there are many diverse components affecting network performance (mainframes, processor interconnection channels, terminals, frontend processors communications processors, and even inter-LAN connections), the performance measurement task is not as reducible and, in fact, is much more composite than a simple microcomputer LAN. This could be a further argument for simple metrics common to perhaps all network components when assessing overall network performance, except when the focused need

was for isolated performance of one component. The integration of network components, however, makes isolated measurements all the more difficult and overall network performance more of a challenge. Further research is needed to determine if there is some combination of performance among network components, such as a linear combination of component performances, which accurately reveals total network performance.

E. HOW FREQUENTLY SHOULD PERFORMANCE EVALUATION BE PERFORMED?

Performance evaluation of computers or networks is an ongoing process if it is to be effective. In many ways it can be viewed much like an attitude toward safety or economy. To be effective it must be practiced.

Performance evaluation should be used during every phase of the life cycle of a system from conceptual design of the workload through reuse analysis of outdated equipment. Basically, the local organization must determine the final answer to how often to evaluate gathered data. A relatively stable period with satisfied users, no new applications anticipated, and some excess capacity may exercise its performance evaluation talent only to keep it ready. On the other hand dissatisfied users, anticipated workload or application increments, and a generally dynamic enviroment experiencing degraded performance may have waited too long to begin preparing for performance evalutation. In military jargon, monitoring for performance data gathering is a necessary continuing activity while performance evaluation is a critical readiness skill constantly, either exercising or preparing to exercise.

F. LIMITATIONS OF CPE PRINCIPLES IN NETWORK PERFORMANCE EVALUATION

Any version of performance evaluation on technical equipment is expensive and should cease when it is no longer efficient. Increasing efficiency is the goal of CPE/NPE, and so it should be examined itself on that basis. One author suggests a performance evaluation group should be disbanded when the cost of operating the group over a six to twelve month period becomes more than the value of savings which the group identifies. Another stopping point is reached when the system is running to everyone's satisfaction and there is no reason to anticipate a need for improvements. However, the term "satisfaction" measured within an organization can be quite subjective and specific even to the subnetwork level. A third possible ending point for performance evaluation is when the system size is optimized and further CPE/NPE efforts only lead to installation size reductions. Small computers and systems profit more by expert programmers than through use of performance evaluation improvement. The reverse has great implications for NPE in SPLICE. A final situation not requiring CPE/NPE is in evaluation of systems such as real time weapons systems, aviation flight controls, nuclear or chemical or life-support monitoring systems where the real-time requirement is 100% effectiveness and efficiency is not an issue. [Ref. 14: pp. 16-17]

There are cautions to beware of in using CPE/NPE. Realize first of all that no one solution can cure all of an installation's problems. Secondly, generalizations about computers and networks, especially in comparisons between computers and between networks, should be viewed with skepticism. One installation's solution may indeed worsen a problem at another installation. Before trying to apply

results of another installation's project or even any particular tool, a careful assessment of the project and its goals should be made. Here is an example of performance evaluation serving as an extension of the organization's strategic goals. Thirdly, the human element must not be ignored. If a "politically" unacceptable solution will be the result of a CPE or NPE study of it or the people who will have to eventually implement it cannot live with that solution, then the suggestion is to abandon that specific effort or find another way. Pushing ahead is only likely to invite failure. [Ref. 14: p. 17]

Acquiring data will be found to be much easier and more acceptable than interpreting it with a purpose according to Abrams [Ref. 15: p. 316]. Clearly defined strategic goals evidenced in performance metric standards, performance evaluation procedures, and trained personnel who understand the goals can sweep aside any resistance to the interpretation process.

Network tools can in many cases, be adapted for use in measuring a system in general. Statistics, queueing theory, software hooks, bit/byte monitoring, modeling principles, etc. can be applied to networks as well as to computers. The key is to know when and where to apply these to measure network performance criteria and not computer performance criteria alone. Network specifics is the subject of the next chapter.

IV. NETWORK PERFORMANCE EMPHASIS

With the foundations for general computer system performance and performance tool and technique use laid, this section suggests considerations basic to performance evaluation and particular to local area networks. The reader is again referred to the glossary in Appendix A for a detailed description of any terms used in the succeeding discussion.

A. GENERAL COMMENTS

Recalling the previously mentioned comments about performance and its dependence upon measurement of appropriately defined simple, unambiguous, and comprehensive quantities relating to users, the reader is reminded of the narrow context of LAN performance evaluation pursued here. In the next chapter the context is further narrowed to SPLICE Lan's. To appreciate the specific context requires some orientation in network definitions. There are two broad categories of networks outlined by Terplan [Ref. 18: p. 61] each with three subdivisions:

1. Switched networks
   a) circuit switched
   b) packet switched
   c) message switched
2. Nonswitched networks
   a) broadcast networks
   b) data processing systems
   c) data base management systems

The types of switched networks are well-described by Rosner [Ref. 19: pp. 27-39]. We can narrow the context by

observing the definition of a local network and its three classes: (1) local area network (LAN), (2) high speed local network, and (3) computerized branch exchange (CBX). Refer to Appendix A in order to discern differences in the three. Each has different technology, physical design, use, advantages/disadvantages, and performance behavior. The emphasis here, of ccurse, is upon packet switched LAN's employing bus architectures since SPLICE LAN's have bus-type topolcgy. SPLICE LAN's in fact have both terminal-to-processor buses and processor-to-processor buses.

A key point is that LAN's are communications networks to which computers, terminals, and other data terminal equipment (DTE) devices are connected in order to satisfy some functional needs at a desired level of performance. The discriminating point in any local network is that the network is a ccmmunications network interconnecting various distributed computing resources. However, both communications and computing resources generally work together in fulfilling the functional needs of users. The concept of translating logical functional needs (or modules) most cften specified in requirements analysis into a design in the form of selected LAN characteristics is a borrowed concept [Ref. 20: p. 3]. Assuming that there are limits to efficiency of operating procedures, the functional characteristics which users require (needs) along with the size and nature of the workload imposed determine the user's choice of LAN characteristics. In turn, this choice of LAN plus the workload nature and size, determine levels of performance. The interrelationship of functional needs, the size and nature of workload, and the choice of LAN characteristics and their determination of levels of performance is illustrated in Figure 4.1

Performance evaluation describes to what level of satisfaction a user's functional needs are fulfilled. For

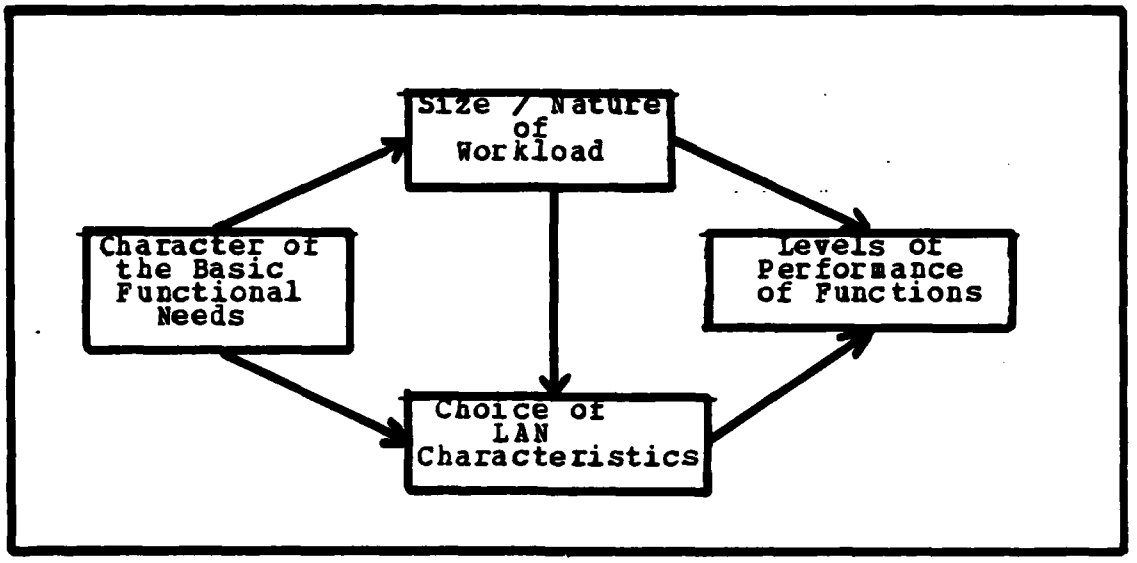


Figure 4.1 Elements Determining Levels of Performance.

this reason it is critical that the performance question of how well the function must be completed be addressed in early user requirements analysis and particularly in the case of networks where there exists multiplicity of interfaces in hardware and software. Once a particular LAN is selected, the levels of performance to be achieved for certain workload demands, such as interactive or high volume traffic, are fairly predetermined. Therefore, some concept of this desired level of performance required to satisfy a functional need must be conceived concurrently with functional needs. This may sound circular, but it is actually in keeping with a soundly established principle of systems design where the outputs (levels of performance of some functions) are defined and designed before all the inputs are (the character of functions, the workload nature and size, the LAN characteristics). As functional needs change or as workload increases or becomes unpredictable, the lack of a performance evaluation effort will deny an organization knowledge of what its network levels of

performance are. Rajaraman supports a similar view of factors affecting performance in LAN's. He says:

> "There are three major factors that affect the performance of the network. They are: (i) the characteristics of the jobs submitted by users, (ii) operational characteristics of the system configuration, and (iii) network interface characteristics." [Ref. 21: p. 34]

Workload accounts for his first characteristic and LAN characteristics encompasses the latter two factors.

B. DIFFERENCES IN COMPUTER AND LAN PERFORMANCE

The differences in computer or computer network performance and local area network performance are not readily apparent if one views all of these simply as systems. The macro view aggravates the ability to distinguish since the tools and techniques applied to computer performance evaluation can likewise be used to assess network performance. The differences can be summarized by thinking about the elements or components functioning in each system. In single computer performance situations the elements interacting, such as the CPU, the input/output channels, the peripheral devices and so on, are generally unique. The uniqueness begins to disappear when the focus is shifted to a computer network where some redundancy of functional components appears as clusters of similar components communicate. Actually, this is not strictly correct since computer networks primarily support communication among the cpu components alone rather than communication among different functional components in separate clusters. This cpu activity is also not restricted to a local area in all cases. In local area networks we see more specialized groupings of resources (groups of cpu's, groups of terminals, groupings of communications subnet

devices, etc.) redundantly spread over the network. Here the specializations observed in the components of a single computer are replicated in a local area network. We see in LAN's the attributes of a high-speed single computer, but distributed with both specialization and redundancy occurring.

A final analogy might help. If one considers a computer as representing a single-celled organism with various functioning components working to sustain the cell, then a computer network might be viewed as a simple colony of multiple similar single-celled organisms functioning together in a symbiotic relationship where some do "batch" jobs some do "data-base" and so on. The local area network analogy is described by a small multi-celled organism where the cells are very specialized and they must communicate through many interfaces to sustain the organism. In addition, consider that SPLICE LAN's do contain a local computer network within them. It is part of the local area network as defined previously.

The impact of this for performance in local area networks is that there are complex interfaces, an often higher volume of activity generated by components of the same functional type, a greater dependence on communications, and a more prevalent occurrence of the human element. These observations support the conclusion that while some performance parameters and behaviors may be common to computers, computer networks, and local area networks, there are behaviors and concerns unique to the LAN's as well. The commonness supports the earlier assertion that CPE tools apply while the uniqueness implies these tools or perhaps others should be applied in other aspects of the entire network. The uniqueness in performance behavior is further narrowed when one chooses a specific LAN to carry out desired functions.

C. LAN CHARACTERISTICS WHICH DETERMINE PERFORMANCE BOUNDS

There are general categorical descriptions of LAN's just as with any system. Beyond procedural and operating adjustments which can affect performance, ultimate style and bounds for perfcrmance are established by which choices within each category are selected for a LAN. One source [Ref. 22: p. 16] classifies local network design issues as either configuration cr protocol ones and visualizes network performance as highly dependent upon each of four basic elements, including transmission medium, a control mechanism, the interfaces, the protocols, and the mutual interaction of these.

Some choices existing among LAN technologies are illustrated in Figures 4.2, 4.3, and 4.4

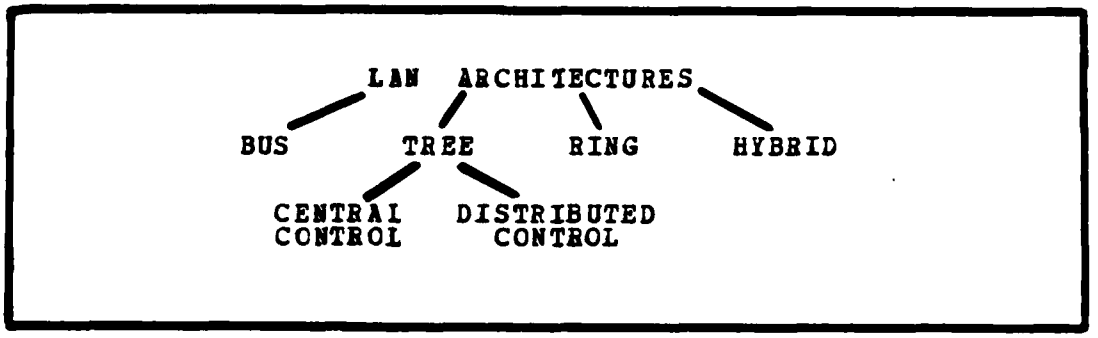


Figure 4.2 Architecture Alternatives.

Terms relevant to this study are defined in Appendix A. The choices exist in the following categories:

1. topology, or architecture (bus, tree, ring, hybrid)
2. access method (dedicated according to time or frequency separation, polling, or random access)

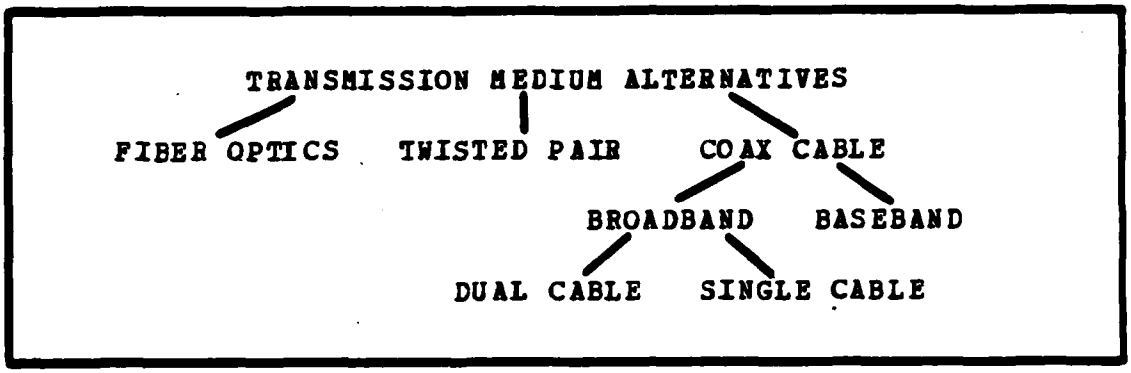


Figure 4.3 Transmission Medium Alternatives.

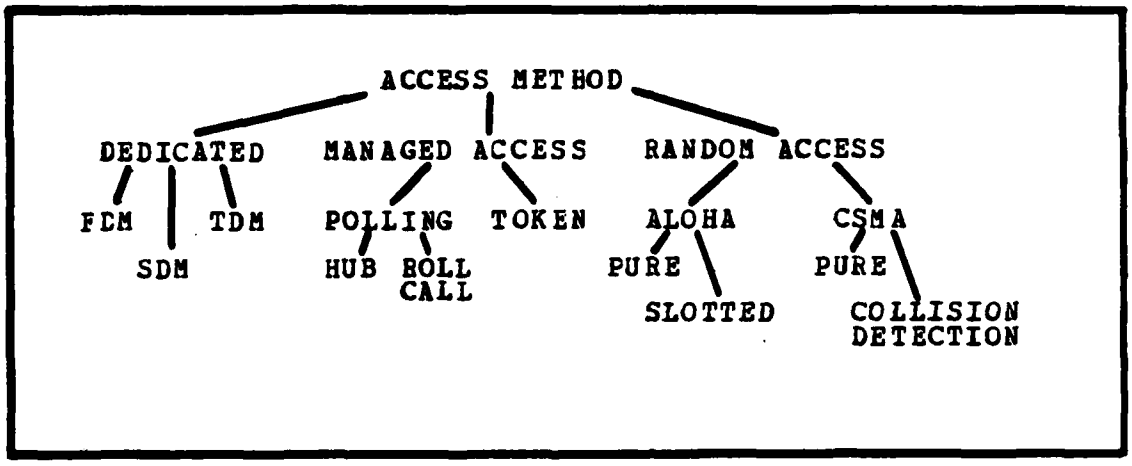


Figure 4.4 Access Method Alternatives.

3. transmission technology (fiber optics, twisted pair, broadband or baseband cable)
4. protocols imposed (low-level, high-level)
5. switching technique (circuit, message, or packet)

Even though these categories are rather independent, some operational groupings of selections from the categories are poor, absurd, or completely unworkable from a cost or performance viewpoint. For example, using a broadband packet switched approach with some sort of polling to

connect a few relatively low data rate devices together would hardly be cost justified. Likewise, to transmit video or integrated digital and analog information over a twisted-pair, random access network would prove disastrous. [Ref. 23: pp. 35-36]

A specific turnaround situation existing with LAN's yet not observed in long-haul networks concerns protocols. Dale Way states:

> "The importance of control (protocol) software . . . (in long-haul slower data rate networks) . . . is minor in determining the throughput performance of the network. The data rate of the link is most often the limiting factor in actual throughput of data from user to computer, or vice versa. The 9600 bps or even 56000 bps rates are slow in comparison to computer rates used in software protocol execution . . . . . now the situation . . . (with LAN's) . . . . is reversed. The link is so fast, the protocols cannot keep up. This makes the health and efficiency of protocol software critical to LAN systems." [Ref. 24: p. 79]

One must realize that performance alone is not the sole motive or consideration in the design of many experimental networks. Such is the case for the Cambridge ring, for instance, which had much more communication bandwidth than initially required, and usage of ring slots was not optimized. More data bit space was available in each slot than was used. [Ref. 25: p. 111]

The selection of a LAN can extend beyond issues of performance or involve tradeoffs in performance as Kee states:

> "Other networking techniques show similar features, with a tradeoff having been made at some stage in their development between cost, ease of implementation, data transmission rate, error rate and intended method of use . . . . . criteria used to evaluate performance could well be different . . . . . the network may be needed to serve a very large population of terminals and personal work stations where raw data transfer speed is unimportant but where a low network delay and the ability to support a large number of users simultaneously are paramount." [Ref. 25: p. 111]

Bailey states:

> ". . . its (a LAN's) construction can be optimized to lower costs . . . . However, not all of the system's performance measurements can be optimized simultaneously." [Ref. 26: pp. 207-208]

Watson, of Lawrence Livermore Labs, points out that in addition to pure LAN characteristics that:

> ". . . the network traffic properties of message size, rate, and distribution have a considerable effect on network performance, and that performance is also very dependent on the mutual interaction of the traffic, the configuration, and the protocols." [Ref. 27: p. 51]

So, we can see that just as the model introduced earlier depicts, the actual performance from a network involves LAN characteristics and more, such as workloads.

D. WORKLOAD CHARACTERIZATION AFFECTS PERFORMANCE BOUNDS

Although trying to characterize and test representative workloads is more appropriately a step conducted in capacity planning, some mention of workload is necessary as it relates to network performance. Specifying workload on a system helps determine which performance parameters should be measured and what trends should be watched.

Workload is simply the total mix of jobs in type, volume, and timeframe imposed upon a system. No doubt if the workload imposed upon a system changes its nature or if the volume of work dramatically increases in a short time period, there will be repercussions for performance. Contention for resources will obviously degrade overall ability of the LAN to perform its intended functions.

Stallings describes the "ideal channel utilization" in a LAN. This description has channel utilization increasing to accommodate any offered load equal to the full system

capacity and then remaining at 100% utilization regardless of further increases in offered load. He notes that any overhead or inefficiency will cause the system to fall short of the ideal. The actual offered load is not the same as the input load of device-generated data put into the network. Actual cffered load can include nct only transmissions, but acknowledgements and retransmissions resulting from errors or collisions. [Ref. 13: p. 235]

Workload is, along with LAN technical characteristics, a contributing factor tc the performance of any LAN. Handling the workload and distributing it among many users is only one of several benefits derived from LAN's.

E. ADVANTAGES AND DISADVANTAGES OF LAN'S

A brief word on the advantages and disadvantages of LAN's is appropriate in discussing performance because if it is the advantages we are attempting to capitalize upon, then those are the very areas which management should be interested in for performance evaluation. Of the five computer-communication problems [Ref. 28: p. 2] commonly recognized, the first three of these are solvable through LAN's. These five commonly observed problems are as follows:

1. The central-computer facility problem occurs where several mainframes and peripheral devices are organized into a coherent set of accessible shared resources; (While this does not meet the definition of a LAN previously referred to here and by other authors, we have to realize there are differences of opinion. It is possible such an arrangement could meet the previous LAN definition if resource sharing occurred and the connectivity was among more than simply cpu to cpu.)
2. The satellite remote-computing problem where there is an interconnection of a wide variety of minicomputer-based equipment and associated peripherals to a central computing facility.
3. The terminal access problem where there is interconnection of an intelligent or unintelligent terminal to a satellite computer or to a mainframe facility.

4. The standard computer-network problem of interconnecting the computing equipment of one organization through some single transparent application-independent computer network, such as a dedicated or private packet switching network, to some other set of computing resources.

5. The internetwork-communications problem links together several independent computer networks via gateway computers, so messages can be exchanged among networks and often through several intermediate networks.

In addition to the basic advantages of resource sharing and resource variety, distance independence, and desire to communicate messages, there are several advantages authors attribute specifically to LAN's. These are the following:

1. System evolution is afforded with impact due to incremental changes under control [Ref. 13: p. 4].

2. Reliability/Availability/Survivability are enhanced with multiple interconnected systems, dispersed functions, and backup capability in the form of element redundancy [Ref. 13: p. 4].

3. With some planning customers do not have to be locked into a single vendor source. [Ref. 13: p. 4].

4. There is improved response/performance in areas of an organization where the service was previously not available or slower than desired [Ref. 13: p. 4].

5. A single terminal can allow a user to tap into multiple systems. [Ref. 13: p. 4].

6. Equipment can be flexibly located. [Ref. 13: p. 4].

7. Integration of services such as data processing and office automation can occur [Ref. 13: p. 4].

8. Fewer data transmission errors than long-haul networks [Ref. 29: p. 52].

9. Significantly lower communications costs than long-haul networks [Ref. 29: p. 52].

The disadvantages are noted by Stallings to exist also [Ref. 13: p. 4] and can largely be attributed to poor planning and subsequent loss of control. Some disadvantages are these:

1. Interoperability of components is not guaranteed, and compatibility factors such as half or full duplexing, asynchronous or synchronous transmissions, data speed, software/operating system/protocol usage, ASCII or EBSIDIC data coding, etc. must be considered.

2. Integrity and security of data must be evaluated where distributed data bases are employed.

3. Difficulty of management in enforcing standards or policies and overall control of data resources is a problem [Ref. 13: p. 4].

With the knowledge that LAN performance can be expected to differ somewhat from that of computers or computer networks, that the selection of a particular LAN capability interacts with workload demands to determine performance, and that advantages of LAN's provide some insight into their performance, we can turn to a discussion of specific network performance parameters.

F. LAN PERFORMANCE PARAMETERS (FOR BUS TOPOLOGIES)

1. General Comments

With the preceding general discussions on standardized performance metrics in mind, a look at details of suggested LAN performance parameters is very appropriate. Traditional performance evaluation has focused upon individual machines. Network performance evaluation centers around message flow in communications links and the overall impact upon traffic in the network [Ref. 21: p. 34]. Rajamaran states the following issues with regard to performance problems in LAN's:

> "Two major issues are important here: (i) the characteristics of the service to users and (ii) the needs of the network management . . . (or network performance evaluation (NPE)) . . . team. Because the users of local networks are mainly within the organization there is a need to satisfy their demands quickly . . . the network management team . . . (and an NPE team) . . . is better able to monitor and take faster action and exercise better control over network resources." [Ref. 21: p. 34]

This dual issue division of network performance coincides with other authors' views:

> "In a packet switched network environment, network performance parameters may be divided into user-oriented and network-oriented performance parameters." [Ref. 30: p. 508]

Stallings as well cites a user's versus a network analyst's view on how information about a network workload may be used:

> "The user may want to know the throughput and delay characteristics as a function of . . . the input load. Or if the network is the focus, the analyst may want to know what the offered load is given the input load." [Ref. 13: p. 235]

While it is inviting to think that all LAN performance parameters could be categorized into various identifiable groupings under two major divisions, "user-oriented" (external metrics) and "network management-oriented" (internal metrics), the many diverse categories researchers have attempted and the lack of standard nomenclature defies any such crisp indexing. The best that can be achieved is to relate some sample logical categories of performance parameters for LAN's and describe the currently defined parameters which particular researchers have indicated belong in each category. The variety in what experts feel are important network performance parameters points to the need for standards as previously argued. It is encouraging that most authors have tried to observe the user versus network performance views. A review of performance parameters, indices, and measurements will be made so that representative ones for SPLICE LAN's can be chosen and discussed subsequently in a separate treatment. Again, the reader is referred to the glossary in Appendix A. Some terms may be described in the text where essential to the discussion.

2. Detailed Performance Parameters

Rajaraman's view [Ref. 21] is chosen first because it may provide the reader a framework from which to think about performance terms and their applicability. This

author supports the dual user and network views and further subdivides these into measurement parameters which can be arrived at through a study of the job flow and the operation of the network. These parameters are then used in various combinations to define four types of performance measures. This author asserts that parameters should provide information about limitations of the networks, should identify bottlenecks, and should be available from data gathered or through calculations upon these data. The major factors affecting network performance are used to categorize the measured parameters. There are three categories, the first of which relates to users and the workload and the other two relate to network characteristics and management concerns.

The categories are as follows:

1. Parameters related to job characteristics (user and workload oriented and determined):
   a) Type of job (whether batch, interactive, multi-user, express, graphics, or device-specific).
   b) Memory requirements of the job.
   c) CPU time requirements of the job.
   d) I/O time requirements of the job.
   e) Job priority.
2. Parameters related to operational characteristics (network-oriented and can usually be set by the operating system or by manual operator intervention):
   a) Parameters for job queue management (affects position of job and progress in queues).
   b) Anticipated field length (identifies amount of memory required by the job before it can be swapped in and is usually different from user memory requirements, but not exceeding it).
   c) Total number of user jobs in the system.
   d) Maximum field length for main and extended memory. (Specified by user at job initiation and its value affects the job's initiation and further progress in the network.)
3. Parameters related to network interface (network-oriented and dependent upon network load, machine availability, and interface traffic):

a) Number of users (batch and interactive).

b) Network and mainframe status.

c) File transfer activity in the network.

d) Network resource availability versus requirements.

Four types of performance specified by Rajaraman [Ref. 21: p. 35] areas each having identifiable indices of performance are then derived from the above measured parameters:

1. System Performance (includes average productive time, average throughput time, job throughput efficiency, average job delay time, and backlog ratio).
2. System Component Utilization Measures (cpu utilization, HYPERchannel utilization, mass storage utilization).
3. System Interface Efficiency Measures (file transfer efficiency and file transfer completion measure).
4. System load (percent of job load by class, abort ratio, and abort time ratio).

Rajaraman [Ref. 21: p. 35] then calculated for each of four performance processors in his system the indices for each of the four performance areas. A composite measure for the network is derived from these figures.

This composite value is time sensitive itself since it reveals a performance measurement at a given time with a given workload and system configuration. Trends should be developed and documented to adequately characterize "typical" performance.

This approach is admirable in its attempt to provide structure, detail, a multifaceted view of network performance, and a composite value; however, realize there are details here pertinent only to some similar networks and some possible measures may have been omitted. For instance, there is an emphasis on processor performance here. HYPERchannel is particular to only some networks and interfacing measures are somewhat slighted. Protocol, terminal, and communications software accesses are not addressed.

Another set of authors previously cited in discussing desirable traits of metrics in general offer an exhaustive standardized attempt to establish topology-independent and topology-dependent metrics which facilitate a comparison between ring and bus networks. They make no distinction between user and network parameters. Their work attempted to relate performance parameters to finite state models of bus and ring networks. That treatment is too extensive for purposes here. Selected definitions are included in the glossary of Appendix A. The discussion will be confined to three categorizations of performance parameters for bus topologies. [Ref. 17: pp. 199-207]

Under the heading of topology-independent parameters Amer and Goel [Ref. 17: p. 198] identify four categories of performance parameters as follows:

1. Time-based metrics measured in convenient time unit increments.
2. Rate-based metrics provide relative measures.
3. Ratio-based metrics involve units of length related to time.
4. Count-base metrics are simply multiplicities or frequencies of occurrences.

These performance parameter categories are found listed in the article along with the topology-dependent metrics. This exhaustive list best represents the metrics which have been defined, and many of them are referred to by other authors as well.

Additional detailed metrics suggested by another set of authors and apparently not duplicated above are found in [Ref. 30: p. 510]. These metrics are not further defined because the authors did not bother to define them and the names suggest the meaning. The parameters mentioned by these authors similar to previously defined parameters include number of data packets sent, number of duplicated data packets, and average packet size.

In reviewing most of the metrics discussed above, a common trait is that most of them are internal performance-related, microscopic in scope, and perhaps not revealing much about service levels. Many of them are possibly hardware configuration dependent. Still they may be of use to network managers who require this detail. Reducing a complex set of measurements into a figure of merit approach might be one way to convert detailed network metrics into service user metrics [Ref. 31: pp. 940, 942].

G. NETWORK SYSTEM PERFORMANCE PARAMETERS

With such detailed but not totally standardized metrics available for LAN performance evaluation and management decision-making, one could easily become bewildered unless quite familiar with computer, network, and system performance evaluation in general and unless looking for one or more of the detailed terms above. The approach can be more manageable and still productive if one concentrates on descriptive measures primarily related to user service needs and to the telecommunications nature of all LAN's. The National Bureau of Standards has done some leading work in attempting to standardize the rating of performance and defining of terminology. Dana Grubb and Ira Cotton of NBS emphasize the following points relevant to packet-switched networks:

> ". . . the user needs a set of performance criteria that encompasses both carrier facilities and data communications hardware and treats them as a single system . . . The nine parameters . . . (criteria for assessing how well a system handles information interchange from a user's viewpoint) . . . do not represent all possible performance criteria, but they are the most essential factors." [Ref. 32: p. 41]

Grubb and Cotton stress the user's interest in only external manifestations of network performance and that the

nine factors are not all independent. Any attempt to improve one factor may cause degradation in others.

Several representative performance metrics which apply especially to network users are defined in Appendix A. These metrics include transfer rate, availability, reliability, accuracy, channel establishment time, network delay or response time, line turnaround delay, and transparency. Availability has often been referred to as the single most significant parameter a user desires Marie Keifer writing for TELECOMMUNICATIONS magazine says this:

> ". . . . multipurpose networks have a better record for downtime. The downtime record actually improves with increases in the size of the network because transmissions can continue temporarily on alternative lines until malfunctioning lines are restored . . . ." [Ref. 34: p. 32]

This assumes, of course, that you have some redundancy of critical lines or components since all networks are not constructed with that in mind. Reliability is as critical for users as availability. Grubb and Cotton [Ref. 35: p. 6-24] describe reliability as a performance metric which describes an aspect of network performance after it has accepted a message from a source for delivery. With regard to response time, Sussenguth cites work done by A. J. Thadhan:

> ". . . . the productivity of interactive terminal users can be improved by a factor of almost two when the response time is reduced from two or three seconds to less than one-half second." [Ref. 36: p. 886]

The line turnaround delay in half-duplex lines is lessened somewhat by transmitting in larger blocks of data, according to Grubb and Cotton [Ref. 35: p. 6-26]. Transparency is listed not so much as a feasible metric as it is an item of great importance to users. Further detail will not be

pursued here. The important thing to note is that studies of these parameters, some of which are pertinent to SPLICE, have already made and could be useful in assessing future performance evaluation of SPLICE networks.

H. OTHER NETWORK PERFORMANCE PARAMETERS

To be sure there are other versions of the performance metric approaches already described. One very interesting idea concerns a universal flow and capacity index as an overall measure of network "efficiency". Of all the research work investigated, this performance measure was the single one which reflected the most comprehensive view of network performance without becoming overcome by details. It addresses the network management orientation more than a user's perspective. The author of this idea summarized:

> "There is no predetermined optimal value of Index I for a network. The purpose of calculating I is to provide a benchmark for adjusting the network so that a subsequent calculation of the index would reflect less interchannel variation. Thus the measurement is a relative one, being most useful when used to compare different networks or new configurations of the same network. For instance, if reconfiguring a network's flow and capacity allocation leads to a lower value of I, then the network is more efficient . . . index I yields a simplified view of a network by tying the multiplicity of its components into a unitary measure that indicates how efficiently these components constitute the whole." [Ref. 37: p. 173]

The final sentence of the above quotation is germane for LAN managers and for future researchers. The reader is encouraged to consult the referenced article for details.

A final source which treats LAN and HSLN performance topics rather thoroughly is Stallings. Regardless of the methods chosen to monitor and measure performance or the metrics chosen to measure, there are three LAN/HSLN regions of operation of which management must be constantly aware. These areas outlined by Stallings [Ref. 13: p. 244] are as follows:

1. A region of low delay through the network where the capacity is more than adequate to handle the load offered.

2. A region of high delay, where the network becomes a bottleneck. In this region, relatively more time is spent controlling access to the network and less in actual data transmission compared to the low-delay region.

3. A region of unbounded delay, where the offered load exceeds the total capacity of the system.

Clearly the third region, saturation, is disastrous and must be prevented. The second region should be avoided through careful planning. And some version of the first region should be a clearly defined strategic goal achieveable through sustained performance evaluation of predefined and standardized network metrics.

I. SELECTION OF PERFORMANCE PARAMETERS IN SPLICE LAN'S

If the avowed system performance requirements of on-line response times and batch processing throughput are taken as the goals [Ref. 4: p. 70], then much of the selection of performance parameters in SPLICE is categorically defined. Such measurements emphasize the importance of communications aspects and user aspects of SPLICE workloads. This does not neglect the importance of details of network measurement since these can easily affect the communications and user qualities of any network. With this in mind the following suggestions are offered for SPLICE performance metric selection:

1. Balance selection of user-oriented parameters with network-oriented ones.

2. Because of the inherent uniqueness of each SPLICE LAN realize that some performance metrics must reflect the local configuration. LAN performance bounds must be apparent in the choice of metrics.

3. Some parameters may depend upon terminal location within the LAN context or upon SPLICE node location and priority within the internetworked SPLICE system.

4. Assuming availability and reliability are regarded as valuable at some defined levels, the indications are for SPLICE that those measurements of time and rate are probably most appropriate and in line with the

response time and throughput goals of interactive and batch processing. Ratio- and count-based metrics while helpful or interesting in a capacity planning sense probably do less for the user and more for those who want to compare systems or parts of systems.

5. In the area of interconnection of SPLICE LAN's with regard to performance metrics, the decision must be made based upon whether more emphasis is desired for network access to other users and network resources, for network services available when LAN and LHN compatibility are achieved, or for protocol functions enhancing internetworking of LAN's via a LHN. (This will be addressed again in a later chapter on internetworking SPLICE.) [Ref. 38: pp. 4-10]

6. Concentration, at least initially, on simply defined and consistently measured metrics will likely pay larger dividends than trying to obtain a measurement for everything or in trying to optimize all performance measurements taken.

7. Regretably in SPLICE's case, the two goals cited may not have the overall network efficiency in mind as universal flow and capacity index mentioned might. The goals in SPLICE appear to be more suboptimization ones which may necessitate causing other performance parameters to be neglected.

V. EVALUATION AND INTERPRETATION OF SPLICE NETWORK PERFORMANCE INFORMATION FOR CAPACITY AND CONFIGURATION PLANNING

A. OVERVIEW

The large investment and operating costs of data communications have caused a heavy emphasis to be placed on the advanced planning function. Assuming that we can adequately gather whatever performance data are desired, there still remains two essential steps to ensure optimum use of that data. These are data reduction, or analysis of the data to appropriately categorize it, and then the interpretation of this analysis in a strategic sense so that decisions affecting the modifications and evolution of the network can be made in concert with organizational needs, policies, and even constraints (budgets, binding contracts, specific mandates, etc.). As previously noted, the first two steps, performance measurement and performance monitoring belong to the Network Control Center (NCC) activity as it carries out the operations portion of network management. The interpretation of this analyzed data requires other portions of the network management responsibility. Those network management responsibilities of planning and configuration management will be concentrated upon here and will finally bring us closer to a position of being able to understand and to make recommendations for strategic network performance evaluation. Assuming knowledge of the workload can be obtained or forecast, we then have some basis from which to structure performance evaluation activity. Studying the research available on subsystems within SPLICE LAN's such as

the TANDEM FEP's, the host mainframes, HYPERchannel, protocols of vendors and of DCA, and terminal characteristics is one way to orient thinking and gain experience in directing performance-related questions. The results of these studies can assist network performance evaluators and capacity planners in interpreting monitored and analyzed data results at SPLICE facilities. Hereafter, strategic performance evaluation management will be referred to as capacity planning. Terplan views CP as one of four parts of overall "network performance management" [Ref. 18: p. 59], while Cortada views system performance management as a separate preceding activity necessary to provide inputs for capacity planning [Ref. 39: p. 56]. The view chosen here parallels Cortada and will be that "evaluation" is the key word differentiating capacity planning from performance management. Dr. Allen of IBM's Information Systems Management Institute cited Richard Armstrong of IBM as saying that performance management is a process of configuring the system to provide satisfactory performance for current workloads and is often called "tuning". While this may not be a day-to-day process, it is usually performed on some discrete frequency basis, and on selected components while capacity planning is a long-term ongoing process of basing decisions to alter network resources upon performance trends interpreted over time. Lynn Hopewell points out the following:

> ". . . long-range planning only makes sense from a total systems standpoint . . . long-range planning must consider so many broad areas of uncertainty that it can only be effectively carried out on an overall systems basis." [Ref. 40: pp. 562, 564]

Hopewell's discussion of three types of planning in telecommunications (short, medium, and long-range) imply that the size, complexity, and interaction of so many

subsystems leaves long-range capacity planning as the only viable alternative.

B. CAPACITY AND CAPACITY PLANNING (CP) IN GENERAL

1. Definitions

James W. Cortada describes "system" capacity as a whole and indicates that it involves measuring user service requirements, availability, workload, and resource ability to handle demands. Dr. Arnold O. Allen quotes N. C. Vince who says:

> "Capacity planning is the means whereby one can achieve meaningful forward estimates of the resources needed, both hardware and software, relative to the demand expected to be imposed by the workload." [Ref. 41: p. 324]

Cortada refers to capacity planning in this way:

> ". . . as a methodology or as various techniques that encompass a set of actions all geared to defining workload characterizations, forecasting workloads, current and future performance, and availability of resources." [Ref. 39: p. 55]

Terplan applies capacity planning to networks as follows:

> "Network capacity planning is the process of determining the optimal network required to meet future service needs. It is based on data on network performance, traffic flow, resource utilization, and estimated growth." [Ref. 18: p. 59]

Note that in this network version of CP, that network performance is a condition preceding the process of CP and upon which CP heavily depends. It is necessary to realize, too, that no system performs at 100% capacity and that capacity of the computer(s) is only a part of the system overall capacity. [Ref. 39: pp. 56, 62]

2. Purpose of Capacity Planning

Capacity planning has many direct and byproduct benefits. Among the direct ones is that it usually fits in with the desire to reduce costs and optimize the network assets. Its overriding objective is to raise utilization of existing resources across the entire system and to determine the need for more. Frequently, 10% of the applications occupy 50% of the resources. If desired this is permissible, but as applications grow in size and number, provision for the smaller applications must be made as well. Capacity planning also prevents panic planning issued in response to crises and resulting in often disastrous decision-making. It primarily assists management in understanding and dealing with change often imposed by a combination of controlable and uncontrolable events. In that light it provides management a means of explaining change to higher authorities in a way which is perceived as more reassuringly under control. [Ref. 39: p. 62]

Byproducts include that CP merges with organizational strategic plans, increases the stability of the system and the organization, and provides the workforce with an element of leadership. This can cast the CP and perhaps the entire DP staff in a more creditable light. Effective CP will accumulate a data base of situations and solutions which can prove of immense aid in helping to reach future semi-structured situations. Lastly, during the operational phase of the life cycle, CP serves to help delay the day when an entire system or subsystem must be replaced, and in the transition phase it gives advance warning of when saturation will be reached. [Ref. 14: p. 14-15]

3. The Capacity Planning Process

a. General Description

Capacity planning (CP) is an iterative, ongoing process if it is to be successful. Terplan [Ref. 18: p. 61] describes capacity planning as consisting of the following five steps:

1. Workload characterization
2. Service-level assessment
3. Workload projection
4. Evaluating network requirements
5. Future network assessment

A network capacity planning methodology presented in Figure Terplan is as depicted by Terplan [Ref. 18: p. 79].

Capacity planning pertains to all portions of the network, including cpu's, data bases, protocols, interfaces, data buses, frontend processors, terminals, operating or control programs, and any other network resource. Good CP involves understanding even non-network resources, such as software or hardware monitors and accounting data packages which require some system overhead to operate, impact upon performance and, therefore, capacity planning.

Seldom does network performance change rapidly; however, management often attempts to implement actions posed by decisions without consulting users and, in some drastic cases [Ref. 39,42: pp. 53, 50], without a plan. CP requires a committment by top management to support it. Participants in CP studies can include a variety of members as well as seasoned experienced staff. DP technical staff, operations personnel, user community representatives application development staffs, and data processing management all have a place in CP.

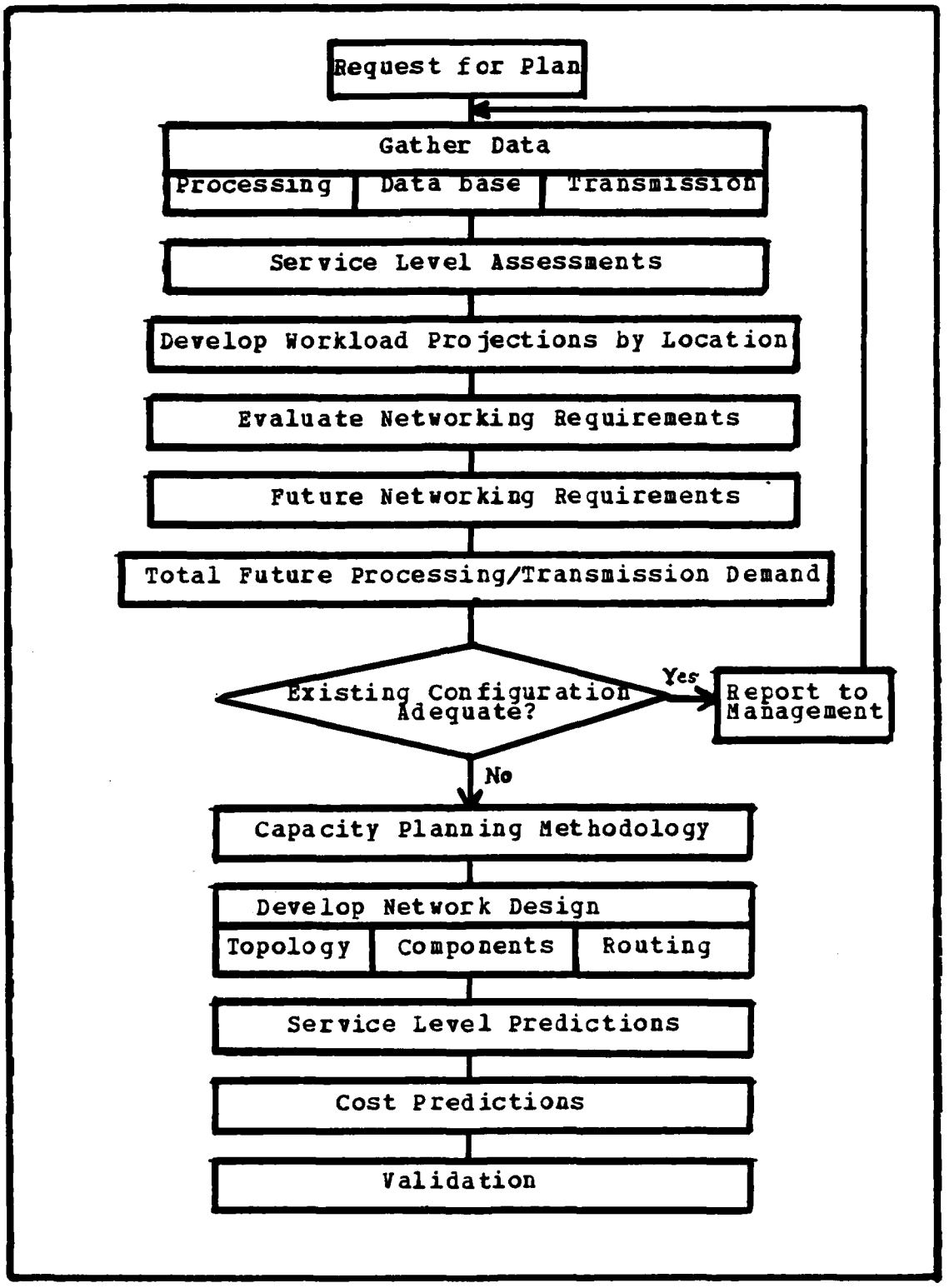


Figure 5.1 Network Capacity Planning Methodology.

The types of questions addressed in CP studies can involve a limitless range of concerns. Examples are these:

1. How much excess capacity should be maintained for absorbing surge capacity or unexpected applications?
2. What will be the impact on response time by adding a certain number of additional terminals?
3. What will be the impact on system performance of adding/modifying an application?
4. To what degree should components be fault tolerant or redundancy engineered?
5. How will performance degrade if a specific data gathering or event measuring package is applied to the system?
6. Is performance different if applying the package at only certain points in the network or at a certain frequency of application?
7. Will adding additional equipment degrade the DP department's capability to service existing users?
8. What are the effects of a new protocol?
9. When should cpu capacity be expanded?
10. Will a configuration change be necessary to accommodate a new technology without degrading service?
11. When will additional peripherals be needed? Additional memory?

The manager engaged in CP generally has one of these choices: (1) take no action to see if his system absorbs the new load, (2) alter some aspect of the hardware configuration, (3) alter some aspect of software, or (4) alter/institute operating procedures. Before selecting an alternative, however, the performance calculation problem stated as follows by J. P. Buzin must be resolved:

> ". . . . given a description of a system's hardware, software, and workloads, determine how the system will operate. Specifically determine throughput, response times, utilizations, and so on." [Ref. 43: p. 347]

Although this was written in a computer-only context, its systems orientation and use of system performance terms

already mentioned by Grubb and Cotton seem appropriate for LAN's. Solving the performance calculation problem involves discerning which tool or combination of tools is appropriate for a particular LAN given the constraints of that tool. Then the tools are applied to predict performance. Interpretations of the results then lead to CP decisions.

As you can see, the workload is a key factor. Much of the information the performance evaluation person or team requires involves external factors which are unknown unless management provides them. This is a critical point of managerial support and involvement in CP. These factors include ones such as number of employees in the organization, organization budget, number of new projects started or applications anticipated, and current and previcus success in meeting the service requirements demanded.

b. Workload Characterization and Evolution

As with tools and techniques of measuring performance, the term "workload" will be borrowed from the CPE wcrld. Workload is simply the mix and frequency of job or resource demands imposed upcn a system and requiring some commitment of network resources. Characterizing the worklcad is the first and a critical step in successful capacity planning. Most research work in describing worklcads has been relative tc computer systems and mostly for existing systems where test generation and sampling of worklcad has been easier than in systems which are being developed or are in planning. Although much of SPLICE's software and hardware is in place with the workload demonstrated in representative benchmark tests, the system is nct yet a complete SPLICE network system, and the actual total local and internetwork load can only be estimated. Strategic planners will tell you that long-range predictions

are generally less precise than short-range ones because information relied upon for long-term decisions is often less accurate and less precise. Accuracy may deviate up to a 50% level for a five-year period [Ref. 44: p. 119]. For this reason the workload today most certainly will not remain static in most organizations. There is no indication that SPLICE LAN's will be any different and most supply demands have yet to shrink.

There are three steps necessary to fully characterize existing workloads: : (1) understand past worklcads, (2) display present workload elements versus resources demanded to get a resulting program/transaction catalcg, and finally (3) correlate business elements (number of items processed, number of tasks required, number of files updated, number of users logged on, etc.) to resource demand (for cpu demand, line time demand, etc.). To understand past workloads requires an analysis of deadline requirements (completion time for jobs, sessions, and transactions), the application cycles (cycle of running application subsystems such as online and batch in SPLICE) daily cycles (sequence of jobs, transactions, and work sessicns by shifts and work centers), and service requirements (availability, accuracy, response time, etc.). The second step can be accomplished in a variety of ways, but it is frequency of resource use and other patterns which are helpful. The third step can serve as a good basis for predicting future workloads. Usage of measurement tools previously mentioned such as hardware, software, and even network monitors now available along with accounting data results, communications software data extraction, and application monitors such as software optimizers can all be useful in characterizing workload by resource demand. [Ref. 18: pp. 61-63]

In contrast, Ferrari characterizes workload by type description instead of by behavior as Terplan has. The description seems less useful for evaluative CP and more useful in measurement, prediction, and comparison of performance studies. He discusses the advantages and disadvantages of his real, synthetic, and artificial worklcads. The synthetic workload is divided into natural and hybrid subsets. The natural synthetic wcrkload is a subset of basic components in the real workload, whereas the hybrid synthetic workload is a mixture of real and constructed components. The natural synthetic load is our familiar benchmark. [Ref. 9: p. 53]

c. Service Level Values

Service level values are constraints in optimizing a network and are based upon standards, requirements, and cost restraints. Service level values percentages or quantities based on service level parameters such as availability of the entire system response time on terminals, turnaround time on batch jobs, and accuracy. Calculation of these can be very subjective, but an example of accuracy in terms cf residual error rate (RER) was given in chapter III. Dr. Allen of IBM says service level determination is the most difficult part of capacity planning and is not done well unless general planning is done well. [Ref. 41: p. 324]

d. Workloads Projection

Workloads projection is also difficult at best. With present workloads, resource demands can be predicted fairly accurately from sensed growth of business elements or units. However, growth of future business elements and thus demand is not as easily predicted for new workloads, and it becomes more difficult the longer into the future the

projection stretches. New workloads can include software extension, software packages, software modifications, software conversion, improvements through application tuning, latent applications (designed and programmed but not yet in production), and new applications. Help from users is critical in accurately predicting new applications. Frequency of execution, pattern of the frequency, and future resource demand expressed in Natural Forecast Units (NFU's) are all necessary. NFU's are the business elements, such as number of employees which can be potentially logged on simultaneously in a given time period. The key is to find a business-related unit that correlates well with a resource demand and to find a way to convert NFU's into resource demand units. An example is conversion of number of employees logged on (NFU) into cpu or access line usage (resource demand). Data for NFU's can be obtained from organization plans, business elements expected in the past, user interviews, records of numbers of application units for certain time periods, and consideration of similarities with resource demands made by other applications. [Ref. 18,44: pp. 64-65, 123-124]

Workload projection must employ some means of categorizing work or jobs just as supply installations have online transactions, batch jobs, and in the future queries from outside installations. In order to be able to determine how an application impacts the LAN in terms of resources usage, we have to classify applications. Mohr defines a major application as one which uses at least 1% of the total system's resources. The largest application can typically consume 20 to 30% of overall system resources, other major applications another 15 to 20% with the second largest using about 10%, five or six applications also qualifying as major applications, system overhead taking as much as 20 to 25%, and two remaining categories, minor

applications and non-application usage, consume the rest.[1] [Ref. 44: p. 120]

This discussion of resource utilization according to application might be particularly relevant to Navy Supply Stock Points and Inventory Control Points where there are several major applications possibly running concurrently. Work in identifying how combinations of applications use network resources may prove fruitful in capacity decisions.

Workloads exhibit changes according to Mohr [Ref. 44: pp. 121-122] in their nature beyond just an association with particular applications. Workloads display an aggregate trend behavior generally in one of four ways:

1. Monotonic increases where workloads grow at a steady rate. These increases reflect increases in user population or in numbers of transactions. This is the most commonly projected workload and clearly applies to the SPLICE LAN situation.

2. Abrupt changes represent sudden changes or discontinuities in workload or resource usage levels almost always caused by external factors. These are the changes of which to beware. They can be caused by installation of additional terminals (resource drain), abbreviated procedures allowing users to be more productive (resource drain or relief), faster hardware (resource relief), or transition of a major system from test stage to full implementation (resource drain).

3. Oscillatory changes are periodic changes in systems workload due to regularly observed variations in the business environment. There are periods of growth and contraction of workloads due to seasonal, operational, or other influences which seem to build and decay. Although no references cite it, this is a prime example of a need for contingency capacity and also typifies some installations, such as at the end of a fiscal year.

4. Random changes represent daily variability in workload caused by random business processes. They also result in random performance measurements which really do not reveal controllable information for planning.

---

[1] These figures offered by Mohr were not substantiated by any particular study or reference.

A graphical view of these workload behaviors is shown in figure 5.2 [Ref. 44: pp. 121-122].

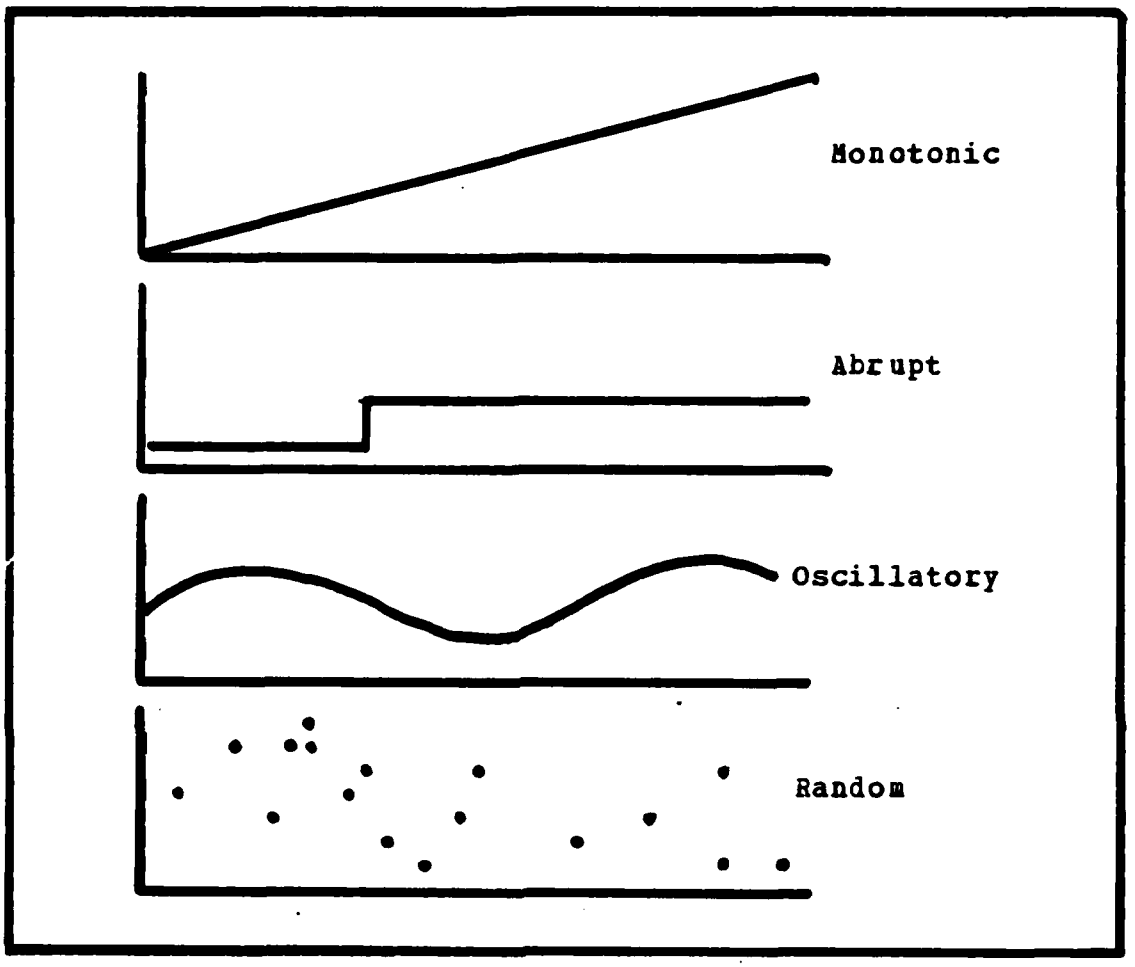


Figure 5.2 Types of Workload Changes.

J. Mohr states in summary:

> "Only monotonic and abrupt changes must be projected. While oscillations and the randomness of the workload must be recognized, they do not represent long-term growth." [Ref. 44: p. 122]

The import for SPLICE here seems to be that most of its workloads fall into the first two categories thus making workload projection a critical part of capacity planning. Some oscillatory and random workloads are a factor and serve as an argument for developing some contingency and surge capacity beyond that for abrupt changes. This is addressed in the SPLICE solicitation document [Ref. 4: pp. C-72, C-73] under "system resiliency". Basically, this calls for a capability to withstand workload variations without interruption to normal loads, and excess capacity of at least 20% with an online response time increase of no more than a factor of 2.0. Provision for returning to a non-saturated condition from a saturated one must be automatic and within five seconds. This latter one may be hard to meet as SPLICE grows without planning ahead for such saturation conditions.

In the workload projection step of capacity planning there is one hidden factor necessitating caution. Many times available performance data reflects an installation's capacity rather than its true workload. The missing parts of the workload are Mohr's [Ref. 44: p. 122] latent workloads. These are the workloads which are not submitted to the system due to some constraint, but if the constraint were removed according to Allen this additional work would appear. This is similar to a subliminal process which is dormant or not possible until the means to satisfy it exists at which time it surfaces. This should sound caution to anyone seeking to project future workloads on the basis of only past history and projections for new applications. Past history data only reflects that capacities are established only to be outgrown and often before anticipated. This only reveals that the workload estimates were in error originally. This latent workload is not to be confused with surge or random events and should be accounted for in capacity planning.

Most CP efforts have used regression analysis and past historical data trends. Since analysis is based upon historical data, any approach using it has deficiencies similar to the approach of simply relying on previous data [Ref. 44: p. 123]. Current trends indicate the use of some type of forecasting unit such as the NFU above to estimate resource demands. Mchr has proposed an approach involving the use of IBM's Business Systems Planning (BSP) and structured analysis techniques to project workloads for new systems. He states:

> ". . . (the approach) . . . provides two major factors that will influence the workload projections. BSP provides a definition of new systems, and the structured analysis provides a description of existing processes and data flows. The workload projection problem then dissolves into processes and volumes of data flows . . . the structured analysis approach can be used as the basis for the projections." [Ref. 44: p. 126]

Some principles of workload projection which are offered by Mohr [Ref. 44: pp. 120-121] include the following:

1. Each major application should be treated individually and detailed projections should eventually be provided for them. Less precision is necessary in workload projection for a new network or application under design. However, the workload should be refined.
2. Workload projection should be at the proper level of precision and appropriate level of detail. That is, major applications which use more network resources must have accurate projection.
3. Since large numbers of minor applications can generally be grouped according to common resource usage, a common workload growth factor should be attainable

e. Evaluating Network Requirements

This step emphasizes the transmission demand while workload projection comments above relate mostly to processing requirements. This demand is caused by growing traffic and addition of new locations. In internetworking

this would mean additional nodes, but in an isolated LAN it implies more terminals or remote job entry locations.

A first step in discovering networking requirements is to use traffic recording (monitoring) machines. These "network analyzers" as they are often called can by measuring carried traffic provide acceptable estimates of network cfferred traffic. This is true at low congestion levels. As congestion increases, additional factors could be measured in addition to the time-, rate-, ratio-, and count-based metrics mentioned in chapter III. Such things as number of call attempts, duration of periods during which no circuits are available, and the number of transactions experiencing congestion (delays or collisions) are additional ways to assist in obtaining an overall picture of the macro network performance indicators. Such macro level indicators include availability, reliability, accuracy, transmission rate, network delay, and so on. This information is logged and then the average traffic is determined for busiest days. Finally, the third step in evaluating network requirements is to associate traffic data with each user's terminal. This may not be trivial to accomplish and assumes that a user summoning an explicit part of an application from a particular terminal can be directly associated with all affected resources. This same problem plagues future resource demand estimation. Future demand should equal present demand plus additional demand expected. However, it is the additional demand forecast for resource usage which causes the real difficulty. [Ref. 18: pp. 66-68]

Before adding new equipment consider two factors: (1) the stimulation factor (almost identical to the latent workload) occurs when more intelligent devices are added which cause an upsurge in traffic, and (2) the controlling factor of providing more control over new

additions to precisely prevent the traffic increase stimulated. This accounts for growth traffic. Traffic at new locations is merely estimated by comparing known sites and traffic profiles with new sites. Beware that new locations can affect the overall network performance merely because of modified rules for routing, procedures for resource allocation, and priorities of service and access. [Ref. 18: p. 68]

f. Future Network Assessment

Future network assessment is a plea to carry out the preceeding four steps of network CP on a continuous basis [Ref. 18: p. 68].

4. Tools and Techniques for Capacity Planning in SPLICE

The issue of which tool(s) or technique(s) to rally behind to help solve the performance calculation problem bears some considerable attention since there is currently no widely disseminated or standardized approach for SPLICE. For the TANDEM suite the results of the original performance tests using twelve representative transaction classes as benchmarks are available. The use of these benchmarks at each major upgrade to the system is encouraged in the SPLICE Strategic Planning Document [Ref. 6: p. 8-3]. While not to be ignored, these benchmarks may now have some drawbacks, not the least of which is that they apply primarily to the TANDEM portion of the network only and they may no longer be "representative" of the response time and throughput performance criteria they were designed to measure. Benchmarks have the additional drawback of deviating significantly in distribution of data across input/output devices when rerun in a real configuration. Load balancing problems differ and performance is not adequately measured. Benchmarks may also not continue to be valid in a capacity

planning situation as the configuration evolves from the originally benchmarked one.

For now, since SPLICE sites are largely in the installation phase of a system's life cycle, Morris and Roth view benchmarks as the primary tool. But moving into the next operational phase (the longest phase) of the life cycle, other tools such as accounting packages, software monitors, and modeling become primary and benchmarks become secondary. Although, once installed benchmarks for a tailored system should certainly be easier and less costly to modify than to develop them for a system in design and procurement. Modeling and benchmarking are most prominent in the procurement phase where the conceptual design of the workload and its eventual specification occur. This specification then leads to equipment requirements where benchmarks are especially useful in preparing requests for proposal (RFP's). Both models and benchmarks provide consistent criteria for proposal reviews of vendor offers. Benchmarks are a virtual checklist to use in the selection process of procurement. [Ref. 14: p. 9]

It is important to note that nonperformance metrics enter into a procurement chcice at this point. Final selection must weigh costs, expansion potential, security, privacy, change and reconfiguration adaptability, operation, technical control capability, manufacturer's support, conversion costs, and delivery schedule tc name just a few [Ref. 17: p. 195].

During the installation phase performance evaluation personnel should draw upon the experience of members of the vendor's service center and conduct thorough diagnostic routines and an acceptance test. This test is conducted to verify that the delivered system's performance is equal to that of the system upon which the benchmarks were demonstrated. The goal is to ensure that the system

installed closely matches the demands of the workload. [Ref. 14: pp. 12-13]

The longest lasting phase is the operations phase of the network system's life cycle. Modeling, accounting data, and software monitors augmented by benchmark reruns and hardware monitor data are called upon to determine the impact of new applications on the existing workload. The objective is to minimize this impact. Program reviews are conducted periodically during this phase to methodically examine the execution characteristics of existing programs and discover those areas where improvement might be possible. New products or applications are best modelled, if possible, as a part of the existing system and workload. This is a cost-effective way to replace components and can assist in helping determine when the entire system needs to be replaced. Adding enhancements has the distinct advantage of delaying the day when the entire system must be replaced. Modeling is usually a good tool to use for predicting future workloads and the saturation point where no amount of enhancements will enable the system to handle the increasing workloads forecast. [Ref. 14: pp. 13-14]

A CPE or NPE team having predicted the saturation point of a system well in advance can begin the examination of new potential equipment and data processing needs in the transition phase of the life cycle. The CPE or NPE people can perform a valuable service also in reuse analysis so that owned portions of a system can be assessed properly for alternative uses and prospective buyers. The life of any system ends as it began with modeling of a conceptual design of the next workload. [Ref. 14: p. 15]

In the area of accounting packages and software monitors SPLICE facilities have a good start. The Burroughs system has an extensive accounting data generating capability now partially used. Standardized procedures for

how it is used and who actually carries out an evaluation or interpretation of the data reduced is not clear. As Morris and Roth [Ref. 14: p. 80], discovered in the software monitor area, most users find software tools easier to work with and their output more relevant

TANDEM Corporation has introduced its XRAY product which has many desirable features although it is heavily and naturally biased to monitoring the system and user processes using the TANDEM equipment within the LAN. XRAY is a software tool for monitoring performance of a TANDEM Nonstop II computer system primarily, although it can be used with other TANDEM software products to measure data base, communications, and even network activity as well. The reader is encouraged to consult the TANDEM literature [Ref. 45: p. 1-1] for a listing of those applications of XRAY for computer and network performance analyzing.

TANDEM claims that through the internetworking TANDEM software product called EXPAND, XRAY is capable of measuring and analyzing an entire network from a single network node system terminal. Features purportedly allow observation of network traffic to, from, and through each node. XRAY interfaces to users via two interactive programs, XRAYCOM and XRAYSCAN. XRAYCOM allows the operator to configure, start, and stop a measurement. XRAYCOM activates a recorder process at each networked SPLICE node processor. The recorder allocates and initializes measurement counters in their respective cpus. The operating system records significant events in the counters, and the recorder periodically copies current counter values into a disc file called the data file. Then the second interactive program, XRAYSCAN, is run on the data file to examine the data in the table or time plot format. This is the data analysis or reduction characteristic of XRAY. XRAYSCAN can be run concurrently while measurement is in

progress, thus allowing the user online analysis of performance. [Ref. 45: p. 1-3]

As covered in Tandem's literature [Ref. 45: p. B-3], XRAY's primary use appears to be for tuning a system by seeking out overutilized components and bottlenecks in parts of the network in an effort tc redistribute workload evenly among available resources, i.e. cause cpus to share the workload evenly, discs to share the workload evenly, etc. Beyond this real time monitoring and operational use, the tables and time plcts can be excellent trend analysis material against which to check workload characteristics, expected service levels, and network requirements. Such software monitor output in short has a relation to CP as well as to current network performance management. Some might contend that when a system is balanced and performance problems persist, a specific resource can be pinpointed as causing the problem, such as balanced input/output limits bounding the performance of a cpu and affecting user response times. The natural and probably correct conclusion might be to buy more input/output hardware. But should the capacity expansion involve more units or should more effort be exerted to enhance capability per unit? It is also quite possible that the software applications themselves require review and improvement in streamlining, thus forestalling hardware purchases, or that operational procedures can be adjusted to alleviate load. Even if hardware is getting less expensive, adding more of it takes up space and adds to the communications effort. Scmetimes there is, of course, no other choice. But remember, CP is an effort aimed at all aspects of optimum resource planning, including hardware, software, procedures, people, and any other modifiable asset. Not simply hardware alone.

Modeling is the tool of choice during Morris and Roth's transition phase. Many difficulties in using

benchmarks are avoided by employing models. However, the model must be valid, the level of detail to include in it must be decided, and the modeling technique (trace-driven, stochastic simulation, or analytic) must be determined. Models, like benchmarks, do require effort and expertise, but with benchmarking share the distinction of having the widest variety of application with respect to a system's life cycle. Network modeling tools are still largely performance analysis and data gathering tools. [Ref. 43: p. 348]

A technique becoming popular is looping. This technique is described in a glossary of Federal Data Corporation's contract award document [Ref. 3: p. I-6]. Looping is a technique of introducing known test jobs, workloads, or diagnostics into a network at a common entry point and monitoring at that site to determine if the expected result returns in a predetermined time and unaltered. Any other response probably indicates some bottleneck requiring isolation.

5. Rules to Observe in Capacity Planning

a) Know the strategic plan of the organization and how CP fits into it.

b) Do CP all the time.

c) Use the correct performance evaluation tool/technique at the appropriate time in the system's life cycle and for the correct reasons.

d) If rules of thumb have been used with success, keep using them and look for others which are not misleading.

e) Use the assistance of your vendors.

f) Know the technologies both inhouse and available in the marketplace.

g) Recognize tradeoffs must exist in any system and 100% utilization of all components is not practical.

h) Recognize the relationship of workload to performance measurement to interpretive CP.

i) Accumulate experience and document it. Future designs may benefit.

j) Select your performance metric parameters carefully and keep them as simple as possible.

k) Increase management's role and involvement in the CP process.

C. PERFORMANCE EVALUATION AND PLANNING FOR COMMUNICATION ELEMENTS OF SPLICE LAN'S

There seems to be more research efforts concerning components within LAN architecture such as the processors, the communications links, and so on rather than overall networks. Vendor technology has, perhaps, accentuated this condition until recently because most products were designed to function as specific standalone equipment or as subunits of a linked group of devices. Networking came along as a concept in combining these components operationally. Only within the last few years have complete LAN's designed from the ground up with separate functionally defined user needs, processing capability, storage and retrieval capability, and communications attributes been available.

Since the future performance of a SPLICE LAN largely depends upon how well its communications subnetwork operates and how it can be adjusted to future demands, a look at the performance of these components might be relevant. We will restrict ourselves to the TANDEM processors, the HYPERchannel connecting the Burroughs mainframes and the TANDEM FEP processors, and the terminal access. Batch processing through the Burroughs while essential to the supply mission, is for SPLICE communications subnet discussion treated as a "black box" which we cannot alter, except through configuration upgrades perhaps. Internetworking issues and implications for performance and capacity planning will be left for a follow-on chapter.

1. TANDEM Nonstop II and Nonstop TXP FEP's

a. General

The fault-tolerant, modular, and independent power source design of the TANDEM processors give them ideal communications function capability and robustness. This robustness is evidenced in the extremely graceful degradation the processors exhibit. Processing continues when components fail, when equipment is being repaired or replaced, and even when new processors or peripherals are being added. They are capable of both multiprogramming and multiprocessing through the GUARDIAN operating system which is entirely duplicated in each processor. As FEP's to operate in the foreground portion of the SPLICE concept, they serve to offload the host processors from teleccmmunications functions and to thereby improve the cost/performance ratio of the system. A vast array of functional, interfacing, and diagnostic software is available, and the vendor's tendency to design modifications and upgrades so that compatibility among units and migration from one generation to a more capable generation is facilitated are positive aspects. The TANDEM systems use all cpus and I/O data paths for processing workloads. No cpu or I/O paths are in a dedicated idle backup mode. This automatically facilitates load-balancing concerns. [Ref. 46: p. 2-2]

Another very positive aspect of the TANDEM product is their overall design to encourage networking. Their experience in this area and a history of satisfied customers speaks favorably. [Ref. 47: pp. 106-107]

b. Processor Performance

As for performance details and capacity considerations, the new TXP 32-bit addressable version

claims to be 20% faster (a 12mHz clock rate resulting in 83.3 ns microinstruction cycle time as opposed to 100 ns time in Nonstop II machines), to provide two to three times greater transaction throughput depending upon the application mix, and to be 2.4 times faster in accessing from main memory than the current Nonstop II. [Ref. 48: pp. 1-4]

The TXP was designed primarily to increase transaction throughput and further optimize on-line transaction processing. What incremental improvement can be achieved by adding TXP processors to a TXP system is not certain, but is probably linear, such that two processors do twice the work of one, four do twice the work of two, and so on. The TXP processor is capable of "pipelining" or instruction overlap to allow concurrent instruction processing in each cpu advantage of faster register-access time as opposed to the slower memory-access time [Ref. 46: pp. 2-4, 2-5].

One clear advantage of the TANDEM system is its built-in redundancy. For instance, the GUARDIAN operating system is redundantly resident in each individual processor and has both "fail-safe" and "fail-soft" capabilites required by the SPLICE functional requirements [Ref. 2: pp. 3 to 15]. That is, a "fail-safe" situation is one backed up by the operating system continuing to direct processing utilizing alternate resources. When alternate resources are not available, "fail-soft" operation is pursued where degraded operations continue. Because all TANDEM cpu's do not share main memory, any cpu failure does not allow such a malfunction to contaminate any memory but its own.

Several Tandem improvements will no doubt impact upon network performance as system upgrades are made. TANDEM's incorporation of the 6100 Communications Subsystem (CSS) with its two dual-ported, programmable I/O Controllers

called Communications Interface Units (CIU's) is a design improvement aimed at removing dependence upon a hardware communications controller. The previous hardware controllers could fail and rewuired manual intervention to select a backup and get the system running again. Data communications movement functions were previously carried out through a hardware controller component and a separate software component residing in the central processor and competing for processing time with other TANDEM software. Now the 6100 CSS serves to offload much of the line protocol management and other teleprocessing control functions from the TANDEM miniconputer software communications processes. As Tandem explains [Ref. 49: pp. 1-2, 1-3], this allows communications processes to attend to their primary job of attending to processing data transfers for the entire LAN. Although shared memory devices might be faster than the systems being placed in operation, the TANDEM system communicating through messages appears to be adequately capable of handling large FEP and query loads on a daily basis despite casualties. This combination of high availability and support for reasonably high response times is an example of a sound subjective management decision fitting the desired performance needs of the organization.

c. Networking Limitations

Limitations foreseen for the processors as a link in the SPLICE communications network are few. Further processor advances may, at some later date, necessitate improvement of the 13 Mbps Dynabus which interconnects TANDEM processors in a cluster. One potential capacity design point is that as the SPLICE environment grows, there are two growth areas assuming a large number of processors might be required at each of several nodes with each node having perhaps several satellite processing sites. The

first is overall SPLICE internetwork growth. The EXPAND software extension of the GUARDIAN operating system and unspecified communications connections can be used to connect up to 255 packet message nodes of 16 processors each (total of 4080 cpu's in a network). This is presumably the wide area network expansion version. This is not to say that other long-haul software protocols could not be used. But use of EXPAND may limit the ability of a node to talk only to other nodes using EXPAND. The other growth area is locally. For intensive high speed processing in a local area network, up to 14 clusters of 16 processors each (total of 224 processors) at no more than 1000 meters between clusters can be connected by the 6700 Fiber Optic Extension (FOX). This is an extension of the Dynabus architecture and provides up to 4 megabytes per second data flow. The network is a ring network. EXPAND software is required for this option also. Long-haul EXPAND nodes and LAN FOX nodes appear no different to a user [Refs. 46,50: pp. 3-8, 3-9, 3-10; 3-1 to 3-5, 4-2, 4-4].

Of course, there is a limit to the traffic a 16-processor per node can handle. Handling all of the internetworking communications FEP duties, local query traffic, and some applications processing may pose a future overload situation. One alternative may be to use processor clusters as described above and to employ one cluster as a dedicated communications cluster for all the other mainframe and job processing minicomputer clusters.

d. The FEP Concept in a Case Experiment

In a final defense of the SPLICE FEP's, one FEP and host processor interface configuration experiment revealed that the central host could normally perform all the telecommunications functions faster than an FEP, but was not necessarily the most cost effective. Using central host

processing power for telecommunications functions is expensive just to achieve less delay in responses. In the experiment the central host was assumed to be twice as fast as the FEP. In tests of four configurations representing different sharing distributions of telecommunications functions between the host and the FEP, the configuration which resulted in the overall least delay was configuration I. In this case the central host was saddled with all telecommunications functions of network control, queueing control, line handling, and editing while the FEP only had to handle I/O transfer of messages. In configurations I and II (where network controlling was added to I/O transfer of messages) the saturation from increasing throughput occurred in the channel indicating that the FEP and host had split the telecommunications duties. In configurations III and IV, the FEP was gradually given the queueing and then the editing functions, thus degrading response times further. The FEP was, in these two configurations, at the saturation point. The ability to modularly add FEP power and grow with the communications subnet workload can be a solution for that situation. [Ref. 51: pp. 215, 216, 227-229]

The design of the experiment suggests an alternative which can be considered for SPLICE nodes when host processing is at a minimum and communications processing is near saturation. This alternative is to find a way to dynamically allow the host to share telecommunications function loads with the FEP. This may more more fully utilize processing capability and delay the need for procuring additional FEP's. Being able to operate near to saturation without actually doing so and still being able to process the workload is a suitable goal.

2. HYPERchannel

a. General

Each SPLICE LAN contains a local computer network as previously indicated. As stated by Carson and Forman:

> ". . . interprocessor communication can apply more stress to a network than can terminal-processor communication." [Ref. 52: p. 92]

Until HYPERchannel, some advances were made in servicing the terminal network, but no similar efforts were made to enhance central site activity so it could keep ahead of burgeoning traffic. Use of a HYPERchannel bus developed by Network Systems Corporation in 1975 has been one approach to dealing with the bottlenecks which developed when trying to locally interconnect heterogeneous hosts, FEP's and storage units in one highspeed local area network configuration (HSLN). This was the first commercially available local computer network architecture. Standard computer channels and I/O control systems just do not have the flexibility to deal with such bottlenecks. The standard channels and I/O control systems were designed from stand alone computer I/O and became insufficient as more data handling devices were attached to a configuration. [Ref. 53: p. 262]

HYPERchannel is a baseband networking product of both hardware and software components operating through a multidrop (up to 64 drop points) coaxial cable and providing for data transfer rates of up to 50 million bits per second. HYPERchannel is a site data channel as opposed to a computer data channel. The coaxial cable has no active elements, and an adapter failure does not affect operation of the trunk. Operational connections are in service up to 3000 feet long, but 1000 feet is more typical.

The original objective with this technology was to off-load local network communications functions from the host as much as possible. This complements the TANDEM FEP concept already discussed. The real key to HYPERchannel performance is buried in the adapters used tc interface various manufacturers' processor and peripheral units to the HYPERchannel. The approach explained by Franta and Heath [Ref. 54: pp. 249-253] was to implement the bottom two layers of a protocol environment in the adapters with four categories of protocols: (1) trunk selection, (2) trunk access, (3) adapter-adapter virtual circuit , and (4) host-adapter, host-host, and host-device This required each adapter to have both memory and intelligence. The heart of the adapter is the microprocessor which consists of a channel interface unique to the attached manufacturer's equipment cn one end, four (expandable to eight) kilobytes of data and one kilobyte of control in a central buffer, and a trunk control logic unit on the HYPERchannel connection side capable of attaching to up to four separate HYPERchannel trunks. Only one trunk is used at a time. The others provide backups and allow additional traffic flow. [Refs. 53,55: pp. 262-264, 50-51]

The first link-level protocol layer allows open and immediate accessibility to the bus for lightly loaded situations and gradually converts to a prioritized ordering of station adapters on the bus as the load increases. Wait flip-flop devices prevent higher priority adapters from dominating the trunk. This trunk access protocol is carrier sense multiple access with prioritized staggered delays for assisting in collision avoidance rather than collision detection as in as in ETHERNET. This protocol is carried through four mechanisms called (1) transmitter disable, (2) fixed delay, (3) n-delay, and (4) end-delay. This protocol is timer derived and is fully distributed throughout the

network. In multitrunk configurations there is also a trunk selection protocol which senses the trunks in succession, searching for a nonbusy one. [Refs. 54,53: pp. 249, 252; 264]

The second level protocol is executed in the adapter's microprocessor where the third type of protocol, virtual circuit establishment, is attempted between two communicationg adapters for the purpose of exchanging a frame sequence. A cpu submits a short request message to its own adapter in an effort to have it reserve its own adapter and initiate a request to reserve the receiving station's adapter for data transmission to follow. When a reservation request is refused, a binary exponential backoff time algorithm activates as with ETHERNET transmission attempts following a collision in that medium. The effect of the transmitting adapter reserving itself is to prevent transmissions from other adapter stations until it releases itself and the receiving station adapter. The fourth protocol type, host-adapter protocols, are simply used by hosts to make function requests of particular adapter staticns. [Refs. 53,54: pp. 249-253, 264]

This is no more than a working overview of HYPERchannel. A more complete treatment of operational details, the protocols, and experimental performance evidence is best obtained in the reference by Franta and Heath.

HYPERchannel is currently the fastest data highway available that has also had extensive research performed on its protocols and general characteristics. It is is quite a bit faster than ETHERNET (50 Mbps compared to 10 Mbps), but ETHERNET is designed for a LAN of up to 100 devices connected over less than a couple of kilometers. HYPERchannel on the other hand links fewer, closer, and higher data rate mini- and mainframes together. This difference makes it difficult to compare them.

b. HYPERchannel Performance

By now quite a few studies have taken place on HYPERchannel performance characteristics; however, many of them have deviated from accurate HYPERchannel operation because of what Franta and Heath call reasons of:

> ". . . mathematical tractability, for lack of understanding of HYPERchannel operation, or because of alterations to HYPERchannel adapters after model development." [Ref. 54: p. 253]

Significant studies of HYPERchannel by Lawrence Livermore National Laboratory and the University of Minnesota confirmed some disturbing thoughts, some of which Stallings summarizes in his text.

Franta and Heath found that performance appeared to degrade when contention occurs resulting in more collisions. This result was qualified in that difference in performance between the enabled and the disabled wait flip-flops (WFF's) diminished as the data length per frame sequence increased or as the number of node pairs increased. A higher throughput was achieved for data frame sequences than for message-only transmissions. These results coincide with other network findings that there is generally less contention and more throughput when transmissions are longer or packet size is larger. Both factors help to prevent the apparent idle medium caused by long propagation delays when two ready stations far apart sense an idle line and who simultaneously then both proceed to transmit with a resulting collision.

There was also less queueing delay for data sequences than for the same throughput of message-only transmissions. The enabled WFF's were actually less fair in allocating trunks among adapters than when the WFF's were disabled. A modification in end delay corrected this in

tests. They determined that throughput, in consonance with Stallings' ideal model, does not deteriorate as load increases. They also discovered that the access scheme does not provide a prioritized trunk access as expected. Interaction of access, trunk selection, and virtual circuit protocols sometimes affected adapter performance by allowing second priority adapters to gain trunk access frequently even at high loads and interferred with the highest priority adapter's ability to use its scheduled transmission time. This is without question the most significant finding of this group for SPLICE managers. If the Burroughs or the TANDEM FEP were the highest priority adapters on the HYPERchannel, performance of the entire network could be affected. [Ref. 54: pp. 253-259]

The Lawrence Livermore group was interested in whether interaction of node placement and contention mechanisms affected performance. While the reader must refer to the referenced article for an accurate description of the assumptions and conditions for the experiments conducted on both HYPERchannel and ETHERNET-like mediums, their results should be of interest to SPLICE managers and capacity planning personnel. Performance was observed to degrade drastically at high loads. This condition was explained by a shortcoming of the level two protocol of HYPERchannel where high loads cause a condition approaching deadlock. The nodes wishing to transmit keep their own adapters out of circulation, and other nodes attempting to establish connections with these nodes do likewise. The result is no one can receive, and adapters are mostly in standby waiting for resources to free up so they can transmit. Deadlock does not occur because postponement periods where a node cannot receive time-out after some maximum wait and the node is returned to circulation. Degradation was also found to be serious when a remote node

was added to the channel. This is not surprising because HYPERchannel is very sensitive to the successive timed sequences in the trunk access protocol. It would follow that an aberrant propagation delay time caused by a remote node would affect performance. They also found that the HYPERchannel contention scheme was superior to the ETHERNET CSMA scheme in terms of stability, or in how well it recovers from an unstable situation of queues developing as a result of an overload. The HYPERchannel at medium to high loads is able to eliminate collisions better than CSMA. The schemes were similar, however, when the remote node and the overload were imposed simultaneously.

3. Terminal Access and Performance in SPLICE

Although a great deal of attention has been given to the central portions of the communications subnetwork in SPLICE, a very user-critical portion is the terminal access for the on-line interacitve users. This subject should not be slighted, but will necessarily be treated briefly here simply because the type of terminal and range of terminal hook-up to the network can be of such great variety. NSC Oakland's case will be cited as an example.

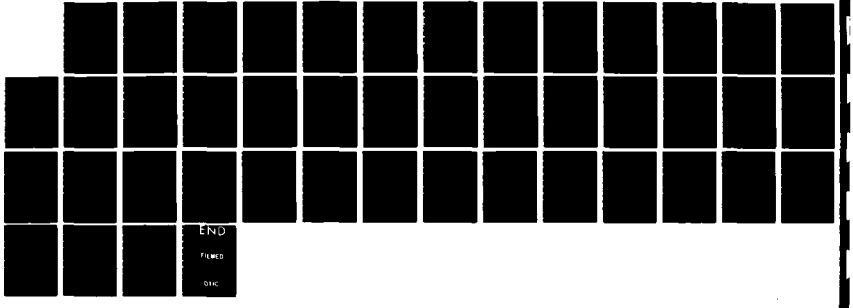
At the current time NSC Oakland has both Burroughs synchronous and asynchronous terminals and TANDEM synchronous terminals. Terminals are arranged six to a shared modem on a coaxial line access from the TANDEM FEP. Federal Data Corporation, one of the contractors in SPLICE, recommended no more than eight terminals ever be connected to such a single drop point. Earlier, FMSO recommendations for using multiplexers to interface terminals to the system ran into operational difficulties during tests. There are 20 to 30 Burroughs asynchronous two-wire direct (TDI) terminals which join the system in a slightly different way. They are point-to-point connected to a terminal concentrator

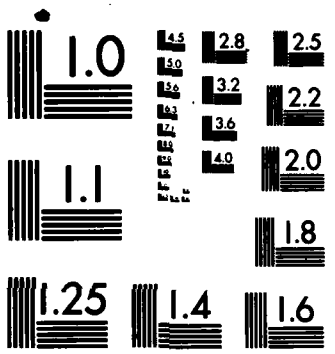
which is connected directly to a B874 miniprocessor used as an FEP for the Burroughs interactive traffic. The possibility still exists for pass-through traffic from 30 synchronous TANDEM terminals and approximately 320 to 330 synchronous Burroughs terminals which access the Burroughs mainframe indirectly via the 6100 CSS subsystem of the TANDEM processors, the TANDEM processors, and the HYPERchannel. The intent has been to gradually move the bulk of the terminals from the Burroughs over to the TANDEM processors as soon as file replication and download of the major applications to the TANDEM from the Burroughs is complete. Despite the fact that transaction ledger on disk (TLOD) files and the recently completed file replication to the TANDEM system for some of the major applications has occurred, the pass-through traffic is still necessary in some cases. [Refs. 56,57,58]

No performance difficulties have yet been encountered, but the system terminals are not yet accessing all applications nor are they attempting to access other SPLICE nodes since the internetworking features of SPLICE have not been implemented . Two possible concerns with the multidrop system could surface as the workload increases with time. First, any connections other than very short bursty ones (such as complicated internetwork ones requiring virtual circuit connection) are likely to cause performance degradation in terminal response times. There is no offered solution to this, however, consideration could be given to prioritizing terminals in certain stock transaction areas and varying the type of hook-up to the system based upon the primary type of traffic handled. An initial short baseline monitoring period to establish traffic loads on each terminal is suggested once significant SPLICE implementation is complete. In this effort may be found some way to trace an individual transaction to discoverr how to individualize

AD-A161 198 STRATEGIC PERFORMANCE MANAGEMENT EVALUATION FOR THE NAVY'S SPLICE LOCAL AREA NETWORKS(U) NAVAL POSTGRADUATE SCHOOL MONTEREY CA D D BLANKENSHIP APR 85 2/2

UNCLASSIFIED F/G 9/2 NL





performance for transactions. According to one source at the operational level [Ref. 57] there is no way to trace a given transaction. Second, the lack of fault-tolerance in this design perhaps for economy reasons makes the terminal access very susceptible to modem difficulties and to single coaxial cable damage. Another opossibility to relieveline contention is to allow terminals to amass blocks of similar transactions to different files in buffers prior to transmission and allow the terminals to sort out a transmission scheme among themselves while the users continue to work.

Conclusions applicable to SPLICE seem to emerge from other studies. One study [Ref. 59: pp. 881-901] done by the Michigan university system's MERIT network directors illustrates the conclusion. Although this is a wide area network example rather than a single LAN or internetworked system as SPLICE will be, it is interesting to note that MERIT terminal usage over a ten-year period steadily increased and gradually occupied more network resources than any other form of processing. The reader is referred to Pawlita's article [Ref. 60: pp. 532, 533] on traffic measurements in data networks for additional possible implications for SPLICE in terminal issues. Some implications Pawlita [Ref. 60: pp. 532, 533] offers include the fcllowing:

a) ". . . medium speed dialog terminals have their own characteristic traffic patterns. (An appropriate question is what is that pattern for Burroughs and TANDEM terminals in SPLICE LAN's?)

b) traffic is extremely bursty . . .

c) uniformly small numbers of input characters, but varying numbers of output charcters . . .

d) strong influence of system applications on terminal perfcrmance . . .

e) different "randomness" of individual user interaction sequences . . . ."

4. Protocols

The subject of protocols has recently been a major stumbling block in SPLICE's progress both in terms of determining which protocols best meet current and anticipated future service needs and due to externally levied SECDEF DDN policy and DDN subscriber requirements. While efforts to ameliorate the situation are showing encouraging signs, other networks no doubt share similar predicaments, and there is more to do to improve protocol performance. One future challenge of measurement and performance personnel alike is that of protocol performance measurement. In SPLICE, for example, there is no way currently to compare the performance of TANDEM's EXPAND software protocol with DCA's Internet protocol without actually running both in side-by-side parallel systems. The results would probably be misleading even if the test were feasible because the two protocols do not necessarily perform exactly the same services. This inability to measure protocol performance and, especially the cooperation of several cooperating layers of protocols, is an aspect of software performance evaluation worthy of investigation [Ref. 60: p. 533].

VI. INTERNETWORKING AS A FACTOR AFFECTING SPLICE PERFORMANCE

A. OVERVIEW

Internetworking SPLICE LAN's for the time being is receiving less priority as each node location attempts to soundly establish its local operations first. Nevertheless, as SPLICE LAN's rapidly come online, there will be growing pressures to consumate the internetworked SPLICE concept.

This chapter is divided into two sections. The first will deal with internetworking issues, including connection of SPLICE LAN's via the DDN. The objective is not to become submerged in technical details many of which are unavailable now anyway. The primary motive will be to suggest those internetworking issues which may affect SPLICE LAN performance. The difficulties lie in the planned decision by NAVSUP [Ref. 2: p. 2-2] to implement internetworking via the TANDEM Corporation vendor protocols, EXPAND and TRANSFER, while the Secretary of Defense (SECDEF) policy [Ref. 61: p. 60] stipulates that all DOD ADP systems and data networks will become Defense Data Network (DDN) subscribers. The latter policy implies that subscribers must use the DDN suite of protocols in order to be fully interoperable with other subscribers and even with other SPLICE LAN's interfaced to DDN. Subscribers wishing to use some other form of long-haul communications must obtain a waiver from SECDEF. SPLICE is currently in this position with NAVSUP [Ref. 62] intending to follow a phased implementation projected to be complete by 1988.

The second part of this chapter will relate ten years of documented experience by a Michigan university system

network. This network is not actually a LAN nor a long-haul network in the sense DDN is, but it does have lessons for SPLICE in the maturing of interactive and batch processing in a network.

B. GENERAL INTERNETWORKING PERFORMANCE ISSUES

1. Protocols and Interconnection

Internetworking of heterogeneous LAN's via a long-haul network (LHN) manifests different performance concerns than merely one LAN with all its components. The connection point of each LAN tc the LHN is a gateway used to connect all hosts in a given LAN to the LHN instead of connecting each LAN host individually. The reader is referred to Stallings' text for LAN characteristics [Ref. 13: pp. 3, 66-69, 74-96], to Rosner's text for LHN characteristics [Ref. 19], and to Schneidewind's [Ref. 38: p. 3] comparison of the features of the two (LAN's and LHN's).

Internetworking involves connecting interface devices such as repeaters, bridges, and gateways. Note that a repeater is an internetworking device for connecting homogeneous LAN's at the physical level and bridges perform similar functions, except with more power and serve to connect LAN's which are not contiguous. Gateways, in contrast, connect noncontiguous heterogeneous LAN's. By such interconnecting of LAN's via a gateway and a LHN, all the advantages of a single LAN are simply multiplied; however, there are prices to pay in terms of tradeoffs in performance, complexity, and costs. In general, LHN's connecting LAN's have slower data rates, higher error rates, and involve distance and routing problems. Complexity is introduced by LHN topology, numbers of subscribers, number of interfaces involved, and the resulting need fcr complex

protocols and the overhead subscribing LAN's experience in handling these protocols in order to communicate with other LAN's. Complexity of interfaces and protocols drives the hardware and software costs. For these and other reasons more than one author has argued for keeping the number of interfaces low and keeping them simple. Sometimes this just is not possible. The need for simple efficient LAN service, the need to internetwork with other LAN's, and the need for compatibility between LAN and LHN at their interface becomes a problem analogous to the one of performance parameter selection in that all needs cannot be optimally satisfied simultaneously. There exist tradeoffs.

The problems arise when the LHN's existence precedes that of the LAN, or vice versa. In the former case the LAN must be designed from the beginning with the protocols in mind which will support the range of functionality desired. If the LAN were designed with no thought to its eventual interconnection with other LAN's, the protocols may be inadequate in necessary generality for internetworking functions even though they may be quite good within the LAN itself. Performance of internetworking processes would suffer later as more interfaces are required either for additional protocols or for protocol conversions. On the other hand, a LAN designed only for the function of connecting hosts to a LHN might sacrifice some LAN performance due to the LHN protocol overhead which will exist in the intra-LAN traffic as well. In either case a second set of protocols generally becomes necessary. It is this protocol proliferation which is undesirable individual and interconnected LAN's. [Ref. 22: p. 32]

The effect of differences in LAN's and LHN's on interface complexity is well illustrated in the following principle:

"The more a local network is designed to increase the effectiveness of intra-local network communication, the more the cost of the interface to a long-distance network increases and the more the effectiveness of inter-local network communication decreases." [Ref. 38: p. 3]

As pointed out by the same author, the same two choices enumerated above exist. That is, the choice is either a LAN tailored to local user needs and a complex, expensive interface to the LHN, or LAN compatibility with the LHN protocols at the expense of some overhead and performance degradation in the LAN. He cites that the tendency is still to design LAN's for long-term local effectiveness and make the one-time sunk high cost for the necessary LAN-LHN interface. The three possible approaches to the interconnection problem cited by Schneidewind reveal a series of tradeoffs. The network access approach emphasizing the ability of a user to access other users and resources implies a need for only the lower three International Standards Organization (ISO) model levels of connection for LAN/LHN compatibility physically and electrically. This approach provides physical connection, but it may fall short of providing all the services needed. The network services approach emphasizes compatibility of the LAN and LHN at higher ISO levels. This approach achieves user services needed, but it may result in intra-LAN performance inefficiencies. The protocol functions approach attempts to please everybody by having one essential set of protocols for the LAN's and another compatible set for the LHN. The result is a need for a complex and often expensive LAN-LHN interface involving protocol translation. In actual fact the author urges deliberate consideration of an appropriate combination of all three approaches to achieve "an effective interconnection". [Ref. 38: pp. 4-6, 10]

2. The Gateway Internetworking Interface

All the issues of internetworking are generally manifest in the gateway between the LAN's and the LHN. Gateways view the connected LAN's simply as "hosts" connected to the LHN.

The burden of internetworking is clearly placed upon the interfacing gateway between the LAN and the connecting LHN. Such an interface, where virtually all traffic flows between (both into and out of) the LAN and the LHN is certain to be a point of congestion. The FEP or other form of connection of the LAN to the LHN also absorbs a share of the load. Clark observes that several LAN's [Ref. 22: p. 33] or satellite LAN's, are connected by bridges in a subnetwork arrangement, the load on the gateway is more severe The connection issues of the lower three ISO model layers, such as speed matching, protocol compatibility, buffering needs, difference in maximum packet size allowable between LAN and LHN and the consequent need for fragmentation of messages, need to maintain virtual circuits when required, flow control, and so on, are all factors which affect the gateway's ability to sort and move traffic efficiently. The more load in terms of volume and in terms of complexity and transformation processes required, the more LAN management can expect internetwork performance to be affected. When a gateway services multiple LAN subscribers to the LHN service, the performance degradation can compound. This degradation causes new problems such as needs for priorities, computing capacity and expense of the gateways, and dissatisfied users.

In short, authorities are seeing the LHN "highway" and its gateway "entrance ramp" as the bottlenecks for internetworking, at least from the viewpoints of LAN users who are getting used to much faster local service. One author states:

> "At this time, it is not clear whether the gateway can assume the entire responsibility for augmenting a local network . . . . with the functions required for communication through a long-haul network." [Ref. 22: p. 33]

Another author writing on internetworking implies that with the two types of packet switching networks, point-to-point networks and packet switching data networks (PSDN), there may be alternatives to relieve the gateway congestion. One alternative is to use point-to-point connections for bulk file transfers where high volume rapid data rates occur even if infrequently. The question here is are the costs of dedicated lines or adherence to a window schedule of usage for several subscribers preferable to the delay in using a PSDN (LHN)? Another alternative and technology challenge is the construction of super multipurpose gateways that can handle the loads and even afford some excess capacity for surges. A final alternative is to build lower-cost specialized gateways or offload some of the gateway functions onto the subscriber LAN hosts or at least to a host FEP. [Ref. 63: pp. 80f, 80j]

C. INTERNETWORKING PERFORMANCE ISSUES FOR SPLICE

All the preceding issues of internetworking and the associated performance characteristics likely to result are issues for SPLICE as well. Rather than burden the reader at this point with more performance metric details, "performance" will now collectively refer to all metrics of interest, such as response delay, availability, throughput, etc. NAVSUP intends [Ref. 1: p. 19] to use DDN as a backbone to serve SPLICE stock point nodes and SPLICE inventory control points (ICP) nodes. The internetworking issues in this plan involve policy as well as technical issues.

As for protocol issues alone, Abrams of the Mitre Corporation states:

> "Growth through interconnections with other networks requires an internetting protocol architecture from the beginning." [Ref. 29: p. 63]

The designers and decision makers associated with SPLICE recognized this and viewed in SPLICE's development that TANDEM's EXPAND internetworking real-time, on-line specialized needs quite well while TANDEM's TRANSFER software could provide time-staged delivery between an origin and one or more receiving sites. There were two complications developing at about the same time this decision was set forth. One was that there were continuing worldwide discussions on protocol standards. The other was that in March 1983 the Secretary of Defense [Ref. 61: p. 1] mandated that all DOD ADP systems and data networks requiring data communications services will be provided long-haul interconnectivity via DDN. [Ref. 61: p. 1]. The subject of protocols was destined to become an issue. The choices are to adopt the DDN protocols and incur the one-time expenses for the conversion or continue to seek waivers and use DDN lines on a closed community basis, but employ TANDEM's EXPAND software for interconnecting instead of TCP/IP. DDN has since adopted the .X25 standard for the lower protocol levels, but insisted that [Refs. 61,64: pp. 10, 22] that TCP/IP is to be used by legitimate (non-waivered) subscribers. The SPLICE operations under a waiver appears to only be postponing the inevitable, setting the possible stage for upheavals in SPLICE when more nodes are further along in implementation, and limiting the variety of other subscribers to DDN which SPLICE sites can success. This latter interoperability issue may not be of concern now, but it can easily become necessary in a

national crisis. The long-term experience base for DDN higher level protocols could begin now. Some consideration for comparing TCP/IP protocol performance with EXPAND performance, if technically feasible, might be advisable. These results and the decision on need for interoperability will no doubt influence the final decision. It is possible that T1 or even more capable lines may come online for DDN and be available for legitimate subscribers to DDN and not at first available to waivered subscribers. For now this can only be conjectured.

The protocol issues are not the only ones which can possibly affect performance in SPLICE. The gateway issues for SPLICE are adequately addressed in Opel's thesis [Ref. 65: pp. 63-83]. The conclusion here is that two half-gateway approach is best for SPLICE. If SPLICE were to adopt the DDN protocols, the protocol conversion from one network's protocol to another network's would be largely avoided. The ICP's currently supported by IBM products have no compatibility with DDN protocols according to NAVSUP, IBM strategic planning does not include such compatibility in the future [Ref. 1: p. 9]. This necessitates using a TANDEM processor to act as gateway (or half gateway) between the ICP's and the DDN. This means more protocols processing, delay, and some measure of performance degradation.

Another issue mentioned in SPLICE Systems Decision Paper III [Ref. 1: p. 16] concerns the capability of DDN to currently handle only asynchronous terminals while Navy logistics terminals at SPLICE stock points and SPLICE ICP's are predominantly synchronous DDN has long-term plans to provide for asynchronous capability, but until it does, once again some measure of performance will likely be sacrificed. This may be viewed only as an opportunity cost since the decision has been made to use DDN and not some synchronous-capable public LHN.

As the subscriber usage of DDN increases, DDN anticipates that SPLICE managements can potentially look forward to added subscribers, new types of traffic, additional DDN nodes, and priority schemes affecting their operations as legitimate subscribers [Refs. 61,64: pp. 5, 7]. This is inevitable since the majority of DDN traffic DDN [Ref. 61: pp. 2, 3] expects to be unclassified and to use the MILNET branch of the unclassified segment as opposed to the ARPANET research branch. Higher priority users will no doubt inflict upon SPLICE performance, but hopefully, on an infrequent basis. Perhaps some supply traffic could be considered of a higher priority than the more routine supply traffic and arguments made for assigning functional sive organizational priorities on DDN. This might require remote contrcls at gateways by the DDN network monitoring stations, prearranged agreements and access procedures to alter priorities, or even real-time communications between the SPLICE LAN sites and the DDN monitoring stations.

DDN casualties and delivery or transmission errors while remote are still possible. The misdelivery error rate is remote at an estimated probability of 5.5×10^{-12} while the probability of an undetected error is even more remote at 4.2×10^{-16}. The analogy for the undetected error rate is that at a steady 24-hour-a-day rate one bit error might be undetected every one million years. Retransmission backup provides for this unlikely occurrence. Should internetworking capability may degrade only slightly. DDN advertises 99.30% availability for single-homed subscribers. SPLICE sites desiring a higher availability can achieve up to 99.99% by dual-homing. That is, obtain more than one access link to DDN [Refs. 61,64: pp. 5, 6; 6, 7].

All internetworking and associated SPLICE performance issues are not tied to DDN-SPLICE factors alone. One of immediate concern is the potential capability of a single

critical SPLICE node to become overwhelmed with query or other traffic. The aggregate SPLICE management must deal with this issue. One built-in control at the current time, at least for single-homed SPLICE nodes, is that the SPLICE LAN can process and handle traffic faster than DDN can deliver it. Unless the local traffic were also heavy or unless a SPLICE LAN had multiple incoming LHN lines, the DDN and its gateway can effectively serve in a pressure-reducing role. Attention to SPLICE node criticality and its placement with respect to DDN access.

A closing issue in internetworking of SPLICE LAN's via DDN concerns coordination. There are many instances when direct communications between SPLICE LAN and DDN authorities may be impossible over the DDN channels. Some form of LAN monitoring site-to-DDN monitoring site communication would be mutually beneficial to both parties and moreso to SPLICE.

D. CASE STUDY OF THE MERIT NETWORK

For an example of network usage statistics reflecting gradual maturity of a network over a ten-year period, the reader is referred to Aupperle's article "Merit's Evolution--Statistically Speaking" in IEEE Transactions on Computers [Ref. 59: pp. 881-902] which describes the Merit network among some of Michigan's universities. This network is not the same in geographic scale, number of nodes, or use of FEP's as anticipated with SPLICE. Merit has 282 terminal parts and only 5 remote batch entry sites. Although Merit is more a wide area net instead of internetworked LAN's as SPLICE will be, some of the same trends and conclusions may be pertinent.

In the Merit network, five network measurement statistics were gathered on a monthly basis over the ten-year period:

1. Number of network users,

2. Successful network connections,

3. Elapsed connection time,

4. Transmitted packets, and

5. Transmitted characters.

There were four types of access to the network mentioned by Aupperle [Ref. 59: pp. 884, 885, 887], including both batch and interactive terminal services:

1. Host-to-host interactive requires a user (or a user program) establish a connection from a local host to a selected remote host. The resources at the remote host that are made available by the network connection are the same as those available through a terminal directly attached to that host.

   a) Classic interactive among two or more network host.

   b) Enhanced interactive where data bases from one host can be used by another host.

   c) File transfer allowing data to be copied from one host to another.

   d) Interprocess communication where programs can run on one or more host computers concurrently.

2. Direct terminal access allowing direct dialing into the netwrok from a terminal and accessing a remote host without going through a local host.

3. External access allows 32 external users to access Merit through an .X25 link and GTE Telenet as opposed to entering via a network host.

4. Network batch service allows a user to submit a job at any network host. Establishing any connections for transmitting the job and for retrieving any output is done by the Network Batch Service.

   a) Remote job entry allows batch job to be submitted either in card deck form or from a terminal.

   b) Batch file transfer allows data to be copied from one host to another.

The overall general picture gained from this study was that the network grew and changed. All curve trends, except the number of megabytes of characters, increased steadily and began a levelling off period only for the last year and a half of the period. Trends such as an initially constant, but later increasing value of packets per connection, a move rapidly increasing value of kilobytes per connection, and a bimodal time per connection are explained as a stabilization

process in the mix of network connect types over time. Batch and terminal access tend to affect these figures oppositely. The number of batch-type connections with short average connect times and large packet- and byte-per-connection transmission rates was observed to decrease while the number of terminal connections with longer connect times and lower data rates increased significantly. These terminals account for the increased connect time trends, while batch work accounts for the increase in kilobytes per connection and packets per connection. [Ref. 59: pp. 889-893]

One notable fact about packet size is that in this study it remained almost constant throughout the ten-year period and under varying operational conditions. [Ref. 59: p. 890]

Specific usage varied by host nodes and according to connection types. The variance in the host nodes usage may hold no significance for SPLICE except to illustrate that each SPLICE node will likewise establish its own usage character. Terminal connection type dominated Merit. Terminal access surpassed external access in average connect times possibly because of cost. This indicates that without some incentives to make good use of terminal time, it will be used as available. This assertion coincides with earlier comments concerning latent workloads and may have implications for user behavior and even necessity of charge-out schemes.

The Merit network has the network attributes of remote access and resource sharing, but is not a load-sharing (distributing a load among the several network computers) or a process-sharing (allowing processes to migrate throughout a network and use resources as needed) network by design. With Merit users must still know where specific resources are in the network, how to access them, and how to use them. This implies a lack of transparency for users in satisfying

their needs. User-scheduled load-sharing, consequently, does not account for much of the network traffic even though the capability is there. This could be a parallel argument for much more transparency in SPLICE to maximize resource use. Most ot Merit's usage increase was attributable to the direct terminal and external services indicating that most user networking needs were rather simple. Analysis of user needs and behavior, if undertaken beforehand, may have predicted these usage trends. [Ref. 59: pp 894, 898, 900]

The Merit study author warns of generalizing too much about Merit data as applicable to other networks. While very little work related to response performance was accomplished in the Merit study, the experience indicated that for interactive computing sessions the network did not introduce noticeable additional response delays. This may be explained by a built-in form of regulation which Merit used and SPLICE management may want to emulate. This is that even though batch and interactive terminal connections were give equal network priority, each host was limited to accepting only one batch connection from each other host while a host could accept interactive connections concurrently. This allowed controlled high data rate batch traffic to flow withcut imposing network response delays. In SPIICE LAN networks where a second and third shift may operate, the contention for connection time may not be a factor and only subject to scheduling.

VII. CONCLUSIONS AND RECOMMENDATIONS

Network performance evaluation and capacity planning are critical elements of any organizational strategic plan and should be intergrated into that plan. Like other organizational elements of the plan, performance evaluation can be approached and viewed as an expression of one way in which the organization can achieve declared strategic objectives.

Capacity planning should be an ongoing continuous effort with flexibility to provide insight into subsystem performance and needs and overall network performance and needs.

Definition and use of some form of standardized, useful, and understandable network performance metrics are suggested. As SPLICE internetworking becomes a reality, the need for standards common to all sites will be manifest. Additional local standards which are necessary for the local area networks or which are local application-dependent may also be required.

Performance was addressed early in the SPLICE procurement phase, but has seemingly taken a backseat in the implementation phase. Before entering the long-term operational phase is a good time to inculcate performance standards and thinking.

At least parttime dedicated personnel assets above and beyond FMSO teams or vendor support is suggested as a vehicle for continuity, and no one better than a resident with the evolutionary observation, documentation, and varied evaluation skills can provide that continuity. Even as much as one person can make a difference.

Management commitment to strategic performance evaluation must exist. Less will only waste the efforts of any network performance evaluation personnel assigned and detract from the credence given by employees to the overall strategic plan for the organization.

Draw upon parallels in computer performance evaluation experience and knowledge and upon vendors for guidance in network performance evaluation.

Adopt DDN protocols to the maximum extent possible.

APPENDIX A

GLOSSARY OF TERMS AND ABBREVIATIONS

This appendix includes selected terms and abbreviations related to the subject of network performance evaluation and referred to elsewhere in the thesis text. The glossary is included as a quick reference for the reader and to prevent the distraction of cumbersome definitions within the text. The pattern of presentation will be to list the term as commonly defined by one or more authors.

1. Accuracy -- "The correctness and completeness of the information accepted by the receiving terminal . . . . Defined in ANSI X3.44 . . . Residual Error Rate (RER) is defined as the ratio of the sum of (1) erroneous information characters accepted by the receiving terminal (Ce), (2) information characters transmitted by the sending terminal configuration but no delivered to the receiving terminal configuration (Cu), and . . . (3) information characters accepted in duplicate by the receiving terminal configuration which were not intended for duplication (Cd) . . . to the total number of information characters contained in the source data (Ct). [Ref. 32: p. 13]

2. Asynchronous -- "A form of communicating where each transmitted character has self-contained beginning and ending indications, so individual characters can be transmitted at arbitrary times." [Ref. 19: p. 355]

3. Availability -- ". . . the proportion of selected time interval during which the information path is capable of performing its assigned data communications function . . . expressed as a percentage." [Ref. 32: p. 43] --". . . the proportion of time when the system is available for use, that is, runs normally. One . . . measure is mean time between failures (MTBF)." [Ref. 10: p. 6] -- ". . . the percentage of the total time during which the system is at the disposal of the users." [Ref. 9: p. 12]

4. Baseband -- "Transmission of signals without modulation . . . digital signals (1's and 0's) are inserted directly onto the cable as voltage pulses. The entire spectrum of the cable is consumed by the signal. This scheme does not allow frequency-division multiplexing." [Ref. 13: p. 351]

5. Bottleneck -- ". . . a limitation of system performance due to the inadequacy of a hardware or software component or of the system's organization . . . The term . . . is sometimes used to indicate the component or part of the system that causes the bottleneck . . . . When the service requests for a given component exceed in frequency and intensity the service capacity of that component, the conditions for the appearance of a bottleneck arise." [Ref. 9: pp. 241-242]

6. Broadband -- ". . . . use of coaxial cable for providing data transfer by means of analog or radio-frequency signals. Digital signals are passed through a modem and transmitted over one of the frequency bands of the cable." [Ref. 13: p. 351]

7. Bridge -- "A device that links two homogeneous packet-switched local networks. It accepts all packets from each network addressed to devices on the other, buffers them, and retransmits them to the other network." [Ref. 13: p. 351]

8. Bus -- "A topology in which stations are attached to a shared transmission medium. The transmission medium is a linear cable; transmissions propagate the length of the medium, and are received by all stations." [Ref. 13: p 352]

9. Capacity -- ". . . the maximum theoretical value that the throughput of a system can reach." [Ref. 9: p 12] -- ". . . quantity of information processing done in a unit of time under a balanced load . . ." [Ref. 10: p. 5] -- ". . . amount of bandwidth originally allocated to a channel." [Ref. 37: p. 171]

10. Carrier Sense Multiple Access (CSMA) -- "A medium access control technique for multiple-access transmission media. A station wishing to transmit first senses the medium and transmits only if the medium is idle." [Ref. 13: p. 352]

11. Carrier Sense Multiple Access with Collision Detection (CSMA/CD) -- "A refinement of CSMA in which a station ceases transmission if it detects a collision." [Ref. 13: p. 352]

12. Channel -- "A path along which signals can be sent . . . connects the message source with the message sink." [Ref. 8: p. 180]

13. Channel Capacity (topology-dependent) -- "The maximum speed of the channel in bits per sec depends on the transmission medium and the electronics at the transmitting/receiving ends). Generally, . . . the theoretical limit as defined by vendor . . ." [Ref. 17: p. 207]

14. Channel Efficiency (ratio-based) -- "The ratio of packet Transmission Intervals to sum of the Packet Transmission Intervals and Packet Transmission Delays. Retransmissions are not included . . . ." [Ref. 17: p. 202]

15. Channel Establishment Time -- ". . . . the time to connect a calling terminal to a called terminal. It includes any dialing mechanism or protocol layer procedures and time required by the network to complete the connection." [Ref. 35: p. 6-25]

16. Channel Idle Interval (time-based) -- ". . . period from end of a Packet Transmission Interval until the first transmission attempt starts not necessarily the time period between transmissions as the transmission . . . may end in collision . . . ." [Ref. 17: p. 199]

17. Channel Length (topology-dependent) -- "The length of the channel from one end to the other." [Ref. 17: p. 207]

18. Circuit Switching -- "A form of switched network that provides an end-to-end path between user endpoints under the control of the network switches. Often called channel switching." [Ref. 19: p. 356] --"A method of communication in which a dedicated communications path is established between two devices through one or more intermediate switching nodes. Unlike packet switching, digital data are sent as a continuous stream of bits. Bandwidth is guaranteed, and delay is limited to propagation time . . . ." [Ref. 13: p. 352]

19. Collision -- "A condition in which two packets are being transmitted over a medium at the same time. Their interference makes both unintelligible." [Ref. 13: p. 352]

20. Collision Count (count-based) -- "The number of collisions a packet of any type encounters before being transmitted." [Ref. 17: p. 206]

21. Computerized Branch Exchange (CBX) -- "A local network based on the digital private branch exchange architecture. Provides an integrated voice/data switching service." [Ref. 13: p. 352]

22. Flow -- ". . . . the throughput as measured on that channel." [Ref. 37: p. 171]

23. Gateway -- "A device that connects two systems, especially if the systems use different protocols. For example, a gateway is needed to connect two independent local networks, or to connect a local network to a long-haul network." [Ref. 13: p. 353] -- ". . . . The gateway may reformat the data as necessary and also may participate in error and flow

control protocols. Used to connect LAN's employing different protocols and to connect LAN's to public data networks." [Ref. 8: p. 190]

24. Host -- "A computer attached to a network providing primarily services such as computation, data base access or special programs . . ." [Ref. 8: p. 191] -- "The collection of hardware and software which attaches to a network and uses that network to provide interprocess communication and user services." [Ref. 13: p. 353]

25. High Speed Local Network (HSLN) -- "A local network designed to provide high throughput between expensive, high-speed devices, such as mainframes and mass storage devices." [Ref. 13: p. 353]

26. Interface -- "1. A shared boundary defined by common physical interconnection characteristics, signal characteristics, and meanings of interchanged signals. 2. A device or equipment making interoperation of two systems possible; . . . . 3. A shared logical boundary between two software components." [Ref. 8: p. 192]

27. Interface Count (count-based) -- "The number of interface connected to a channel." [Ref. 17: 207"]'p.

28. Interface to Interface Communication Delay (time-based) -- "The time from when a packet is ready to be transmitted at a sender interface until the packet has been communicated to the receiver interface." [Ref. 17: p. 199]

29. Internetworking -- "Communication among devices across multiple networks." [Ref. 13: p. 354]

30. Line turnaround delay -- ". . . the time required by half-duplex circuits to reverse the direction of transmission." (Full duplex lines have permanent virtual links and no such turnaround delay. Transmitting in larger blocks of data can lessen this parameter's effect.) [Ref. 35: p. 6-26]

31. Load Balancing -- "A system is balanced when its workload is evenly distributed among all of the available resources." [Ref. 45: pp. B-1, B-3]

32. Local Area Networks -- "A general-purpose local network that can serve a variety of devices." [Ref. 13: p. 354]

33. Loopback test -- "A test in which signals are looped from a test center through a data set or loopback switch and back to the test center for measurement." [Ref. 8: p. 194]

34. Maximum packet length (topology-dependent) -- "The maximum length of a packet that can be transmitted/received over the channel by an interface limited by software as well as hardware considerations." [Ref. 17: p. 207]

35. Message switching -- "A switching technique using a message store and forward system. No dedicated path is established . . . each message contains a destination address and is passed from source to destination through intermediate nodes. At each node, the entire message is received, stored briefly, and then passed on to the next node." [Ref. 13: p. 354]

36. Network Delay -- ". . . the time required for a message to be transmitted from a source and accepted at the designated sink (destination)." [Ref. 35: p. 6-25]

37. Network Power (ratio-based) -- "The ratio throughput to average Station-to-Station Packet Delay . . . reflects how fair a network is to different users." [Ref. 17: p. 202]

38. Offered Channel Traffic -- "At any instant, the total number of packets in the interfaces waiting to be transmitted. The packet that is being transmitted at that instant is not counted . . . This metric depends on the fuffers at the interface." [Ref. 17: p. 206]

39. Offered Load -- ". . . the total number of packets offered to the network." (Denoted by the letter "G".) [Ref. 13: p. 235]

40. Packet -- "A group of bits that includes data plus source and destination addresses." [Ref. 13: p. 355]

41. Packet Switching -- "A method of transmitting messages through a communications network, in which long messages are subdivided into short packets. Packets are then transmitted as in message switching." [Ref. 13: p. 355]

42. Packet Transmission Count (count-based) -- "The number of times a packet is transmitted (original plus duplicate transmissions) before it is communicated. Redundant transmissions are not included." [Ref. 17: p. 206]

43. Packet Transmission Delay (time-based) -- "The time from when a packet is ready to be transmitted in an interface until the start of transmission." [Ref. 17: p.200]

44. Packet Transmission Interval (time-based) -- "The time from when a transmission begins on a channel

until a packet has been fully transmitted." [Ref. 17: p. 200]

45. Protocol -- "A set of rules governing the exchange of data between two entities." [Ref. 13: p. 355]

46. Relative Network Throughput (ratio-based) -- "For the same Offered Channel Traffic, the ratio of Throughput of network 1 to Throughput of network 2." [Ref. 17: p. 203]

47. Reliability -- ". . . . the likelihood that a telecommunications facility will remain operational until until the information transfer has been successfully completed . . . describes the performance of a system after it has accepted a message from a source for delivery." [Ref. 35: p.6-24]

48. Response Time -- (same as network delay) [Ref. 35: p. 6-26] -- ". . . the time interval between the instant the inputting of a command to an interactive system terminates and the instant the corresponding reply begins to appear at the terminal." [Ref. 9: p. 11] -- ". . . the time that the operator must wait to begin a transaction after completing the previous one." [Ref. 16: p. 2]

49. Stability (time-based) -- "If the number of transmitting interfaces (and . . . stations) . . . is allowed to increase without bound, then a channel is . . . . stable if the station to Station Delay stays within Xms, where X may depend on the number of interfaces . . . . . Throughput must be a nondecreasing function of offered channel traffic for the channel to remain stable." [Ref. 17: p. 198]

50. Station to Station Message Delay (time-based) -- "The time from when a message originates at a station until the message is assembled successfully at the receiver station." [Ref. 17: p. 201]

51. Station to Station Packet Delay (time-based) -- "The time from when a packet originates at a station until that packet is received at the destination station." [Ref. 17: p. 201]

52. Synchronous -- "A form of communications where characters or bits are sent in a continuous stream, with the beginning of one contiguous with the end of the preceding one . . . . requires the receiver to maintain synchronism to a master timing signal." [Ref. 19: p.301]

53. Throughput -- "The number of packets communicated on the channel per unit time." [Ref. 17: p. 204] -- ". . . its value may be expressed in many ways: . . . number of transactions processed per unit of time, ..." [Ref. 9: p. 12]

54. Throughput Law -- ". . . system throughput is equal to the utilization of cnly device, divided by the demand for that device." [Ref. 45: p. 5-3]

55. Topology -- "The structure, consisting of paths and switched, that provides the communications interconnection among nodes of a network." [Ref. 13: p. 356]

56. Transfer rate -- ". . . the rate of the number of information bits accepted by the receiving terminal configuration during a single information transfer phase . . . to the duration of the infcrmation transfer phase." [Ref. 35: p. 6-22]

57. Transmission medium -- "The physical path between transmitters and receivers in a communications network." [Ref. 13: p. 357]

58. Transparency -- "In data communications, the ability to transmit arbitrary information, including control characters which will be received as data." [Ref. 8: p. 204] -- ". . . describes the absence of code or procedural constraints imposed on the information processing by the communications system." [Ref. 35: p. 6-26]

59. User Channel Throughput (rate-based) -- "The total number of bytes in all transmissions from an interface per second . . . includes synchronization and check-sum bytes. Bytes . . . involved in collisions are not counted . . ." [Ref. 17: p. 205]

60. User Channel Utilization (ratio-based) -- "The ratio of User Channel Throughput and Channel Capacity." [Ref. 17: p.203]

61. User Information Throughput (rate-based) -- "The total number of information bytes communicated from a station per second." [Ref. 17: p. 205]

62. User Information Utilization (ratio-based) -- "The ratio of User Information Throughput tc Channel Capacity." [Ref. 17: p. 203]

APPENDIX B
COMPUTER PERFORMANCE EVALUATION TOOLS

A. THE "VIRTUAL" TOOLS

Two of the most important and effective tools for evaluation of performance available to nearly every computer installation and often overlooked are the simple ones of (1) visual inspection and (2) common sense. Together they merely compose the essential ingredient of any effective evaluation effort: reflective observation. Morris and Roth [Ref. 14: p. 6] note that any performance evaluation effort starts with a visual inspection of a suspected problem area and is followed by a common sense application of some more specific performance evaluation tool. These two tools could be argued to fall in either or both categories of CPE tools.

B. ACCOUNTING DATA REDUCTION PACKAGES

Perhaps the earliest CPE tool evolving from the use of check flags and counters in the programs of early computers is the broadly used (3) accounting data reduction program. This tcol belongs in the measurement category. Continually more refined versions of these data gathering programs were developed by computer manufacturers or as separate commercial developments. These programs showed an evolution parallel to that of users' needs which moved from check flags, to manual logging and billing, to automated trace routines, and finally to comprehensive data collection programs. Such programs are for the purpose of describing the amount of computer resources consumed by or in support of each application program run on a system and are generally used for billing computer users in some sort of a

charged-out system. These programs are a rich source of information for most performance improvement projects and could be used to document trend usage in support of capacity planning decisions. Some version of these programs is nearly always included in a procurement package and considered somewhat "free". [Ref. 14: pp. 2-3]

Accounting data packages do have some limitations. Very few if any such single package can provide data in every combination and about every parameter desired. They use computing resources in proportion to the amount of work they perform. When such packages are used only for sampling of data for performance and management studies, from 2 to 5% overhead is imposed on the system. If features of the package are engaged, however, the overhead can range upwards to as high as 30% or more. The typical overhead level for a comprehensive package used for CPE purposes is around 10%. Accounting data packages are not for serial-only computers where a data collection routine along with an application program would pose a severe processing burden. However, multiprogramming environments (batch, teleprocessing, and mixed batch-teleprocessing) can benefit in varying degrees. Accounting data systems are best used with batch systems because of consistent batch system behavior which is primarily computer-oriented. Such packages are difficult to use with teleprocessing systems because of the influence of the unpredictable human user element and the decreased visibility of teleprocessing activity. Much of teleprocessing activity is simply generated by software and hardware which is outside the confines of the computer(s) having the resident accounting data package. Another difficulty with teleprocessing systems and accounting data gathering is the need to time stamp gathered data to note when resource usage occurred unlike the mere data gathering in simple CPE systems. Additional code added to perform

this timing notation is more overhead unneeded when monitoring activity is heaviest. Another tradeoff is that systems with less comprehensive data gathering packages generally require augmenting software or hardware monitoring. [Ref. 14: pp. 58-60, 72].

The advantages and experience with accounting data continue to make them a more comfortable approach for many organizations. The length of experience, familiarity, and influence of vendors accounts for the reliance of many CPE teams upon accounting data packages. Such data is considered [Ref. 14: p. 69]. representative, acceptable, and available. Data reports from such packages are widely used by installation managers, programmers, and CPE groups as well. This breadth of exposure is not quite so easily facilitated with other performance evaluation tools or techniques where expertise must usually intervene to produce interpretable results. Another positive sign for teleprocessing environments such as SPLICE specifically is that the next level of sophistication in accounting data packages above comprehensive packages is being perfected. This is the trace or trace-driven system, where noncontinuous tracing, or sampling, of data is done in reasonably short time periods. Such data sampled in an interactive environment could be the types of inquiries or updates made by a user terminal or cluster of terminals in a short time trace or even the user(s) demand for various hardware or other resources in a similarly short time period. Of course, it would again be for management to determine the length of such a period. [Ref. 14: p. 61]

Perhaps the most attractive advantage of accounting data is that it can point to areas where another tool can be used to narrow in on a problem, such as to identify target programs or components for examination by monitors, to tailor simulation inputs, or to characterize workloads for

benchmarking. These other uses will become apparent as the other tools are discussed.

C. SOFTWARE MONITORS

A fourth CPE tool is (4) software monitors. These are also measurement type tools. They are specialized sets of software code integrated into the computer's operating system and used to collect statistical information about the distribution of activity caused by execution of any particular application programs or routines or about use of all or parts of the hardware configuration by the software.

Software monitors are event-driven, time-driven, or a combination, and sampling techniques are used to control their operation. Event-driven monitors work by means of hooks or changes of state. Hooks are recognizable instructions inserted into the operating or control program to cause a set of data to be gathered whenever the hook is encountered. A change of state occurs whenever one type of computer activity stops and another begins. Hooks and changes of state are the events that cause the monitor to operate according to some specified sampling frequency. Time-driven monitors examine a particular activity and collect a predefined data set by using a clock to interrupt processing at fixed intervals. Most successful monitors use a combination. Time-driven techniques are used for frequent short-lived activities and event-driven for less frequent longer events. [Ref. 14: pp. 76, 78]

There are three categories of monitors which include optimizers as well to be discussed shortly. These are Application Program Analyzers (APA's), Control Program Analyzers (CPA's), and Equipment Usage Analyzers (EUA's). EUA's are most like accounting data packages since they gather data on amount and distribution of work for various

system components of a configuration program by program and as a complete system. EUA's simply can get a greater level of detail. These tools differ from acccunting data reduction programs in that software monitors can collect a finer level of detail by examining step-by-step execution of coded instructions. Like accounting packages, software monitcrs are commercially available, but primarily only for a narrow range of manufacturers and mostly for large mainframes. Software monitors are very system-dependent. Since these tools are incorporated into the operating system, some contend that in seeking the performance of an application there is no resulting overhead. Others disagree that any additional software is overhead. [Ref. 14: p. 78]

A brief note on strengths and limitations of software monitors includes the software optimizers since both are programs. The advantages of software monitors include that as prcgrams they are easy to install and use, that they are relatively inexpensive, that they can collect unusually detailed information, and that the commercial varieties come normally with maintenance support experience of a vendor and may have other features such as special reports. Limitations include that they consume computer resources, may produce misleading results when a sample is not large enough, are system and language dependent, and can collect only information accessible through software instructions. [Ref. 14: pp. 89-92]

There have been cases of user-developed monitors causing nearly 100% overhead. One survey [Ref. 14: p. 79] reported, though, that users were more satisfied with software monitcrs than any other CPE tool.

Most established CPE groups seldom find a need for more than accounting data and periodic software monitoring. any additional software becomes overhead.

D. PROGRAM OPTIMIZERS

A subset of accounting packages and software monitors and likewise falling into the measurement category of CPE tools is the (5) program optimizer. These are specialized sets of code usually written in the language of the program to be optimized and compiled with the application program to collect information on execution characteristics of only that particular program when it is run with test data. Program "optimizer" is a slight misnomer because these code sets do not optimize programs. Rather they produce reports that indicate to programmers what parts of application programs might be improved to decrease running time or computer resource usage. Optimizers, unlike accounting data packages and software monitors, can collect information such as parts of a program which are not used or are seldom used. These tools can assist in pinpointing efficiency. Since they are compiled with the application program, they are compiler dependent while accounting packages and software monitors are more system dependent. These tools also impose some overhead upon the system. Program optimizers are primarily event-driven. Their strengths and limitations revolve around their nature as programs discussed above under software monitors. [Ref. 14: p. 4]

E. HARDWARE MONITORS

A tool which is more difficult to use because the user must be familiar with the architectural details of the system to be monitored and because of the voluminous data it can produce is the (6) hardware monitor. This equipment is more of a traditional measurement category tool since it is a piece of electronic equipment attached to the internal circuitry of the system to be monitored for sensing changes of state at these connection points. Information is

recorded or displayed on the number and duration of events occurring at each connection point. The information is saved for later reduction by a specialized software program. Such hardware monitors are called basic monitors and are system independent as long as the connection points of interest are known for a particular brand of computer or network equipment. Mapping monitors incorporate memories and special register adapters to enlarge the monitor capability for simultaneous measurement of large numbers of signals. Reports are produced which cover many combinations of the physically monitored signals that seem like larger numbers of basic signals. These monitors are also system independent, but require a much more detailed knowledge of the systems architecture monitored. The most recent evolution has been intelligent monitors that communicate with the programs executing within the computer to control the information collected by the monitor. These monitors are system dependent, and the monitor must virtually reproduce the monitored system's architecture so operations can be recognized as they occur. Generally, these tools are rather passive and truly monitor without perturbing the device monitored. [Ref. 14: pp. 4-6]

Hardware monitors are in general not for CPE beginners. They require a great deal of systems knowledge, training and practice, and an understanding of the nature of the workload on the system for results to have any meaning. These tools are usually a last resort, but can be productively used in the hands of skilled technicians, especially when the information to be obtained is invisible to a software monitor. [Ref. 14: p. 113]

Strengths of hardware monitors are not as significant as limitations. First, there is no way to correlate data collected with specific programs executed. Second, some control program functions often cannot be tracked. Third,

connection of such a monitor sometimes involves proprietary permission of the computer manufacturer. Fourth, training and experience are a must. Fifth, connecting such a "black box" can be a disrupting ordeal. Lastly, the costs do not stop at leasing, renting, or purchasing the monitor. The data must be reduced and lots of time is absorbed.

F. BENCHMARKS

As previously mentioned, the use of (7) benchmarks as a tool was a primary means of evaluating the hardware and software combinations of vendors competing for the SPLICE project. A benchmark is the term implying a standard for comparison or a point of reference for other products or activities similar to the one chosen to serve as the benchmark. Benchmarks in the computing and network sense are programs or sets of programs used to represent a real workload in operation on an existing computer system or a workload planned to be in operation on an existing or proposed system. Benchmarks are useful for validating or verifying the results of other CPE tools. Benchmarks are difficult to classify as measurement or predictive tools because they have characteristics of both. They are measurement in the sense that they require a system to exist, and they are predictive in the sense that they are used to estimate the future impact of a present decision. [Ref. 14: p. 6]

In this light benchmarks can be described as a strategic tool for determining if a system fits the established objectives of the organization. However, benchmarks are used to validate the impact of operational or procedural changes as well as in procurement situations such as SPLICE.

Benchmarks have the advantages of thoroughness, of more prediction than any of the other methods, and of encouraging

a common criteria or standards approach to performance evaluation. They also have the disadvantages of high cost, of being very time consuming activities, of the requirement for portable software which can be taken off one system and put onto another, of the requirement that benchmarks must be accurate representations of workload, and various external factors. The external factors include the need for human intervention, the occurrence of program bugs, and the possibility of equipment failure during a run. [Ref. 14: pp. 132-133]

Despite the efforts of FMSC and contractors to benchmark the TANDEM systems, one author feels online systems do not lend themselves well to benchmarking. Cortada states:

> ". . . they are easiest to do with batch loads, but nearly impossible with online systems." [Ref. 39: pp. 79-80]

It remains to be seen when the SPLICE LAN's are fully operational if the online benchmarked SPLICE results were an adequate estimate of real workload. Overhead is not an issue with this tool because any alleged overhead is actually some aspect of the test workload benchmark itself.

G. SIMULATION

Another aid to performance evaluators is actually a technique rather than a tool. The technique of (8) simulation does not require the existence of a system for making direct measurements. Simulation uses logical models of a system, concept, or operation to examine its behavior over time. The purpose is to estimate what the measurements would be if the simulated system were to be measured directly. If the simulated system does exist, actual measurements can be used to improve simulation models and

results. The models mentioned are programs executed on "host" computers which are computer systems other than the one being simulated which is the "target" system. Simulations are used to obtain experimental data for insight into a system. Simulation is normally used in conjunction with other CPE tools and techniques. Simulation is most useful when the system is in the design phase, is not installed, is not available, when other tools are not available or cannot be used, and when analytical models are insufficient. [Ref. 14: pp. 135, 136, 138, 140]

Major advantages of simulation include that it can be used with large, complex, and difficult problems. It enables management to make decisions easier by revealing important elements of a problem along with alternative solutions. It is a limited technique in that it is expensive, time-consuming, and can result in misleading results if the models are not validated thoroughly. In these aspects it highly resembles benchmarking. Simulation also introduces personnel problems since experienced simulation personnel are creative and independent and even difficult to manage in addition to the isolated "ivory tower" image co-workers ascribe to them. [Ref. 14: p. 140]

H. MODELING

The last tool to be discussed is (9) modeling. Modeling is the creation and exercise of mathematical descriptions (models) of portions of the system as it should operate if implemented. It is very similar to the simulation definition minus the "over time". Analytical models are sets of mathematical equations whose independent variables (inputs) produce a single set of dependent variables (outputs). The main difference in the two is that analytic models are deterministic where the same inputs will produce

repeatable outputs while a simulation is nondeterministic and produces a range of results or outputs for any set of inputs. Modeling is definitely a predictive tool and is often considered a subset of simulation. Modeling is, however, a discipline in its own right, and the computer field like others has its own specific modeling tools. In CPE the tools are computer program packages that model computer systems. There are also computer modeling languages. Language tools are used when more detailed short time span problems are studied, and computer program package tools are used when overall systems activities amounting up to an hour or more are under examination. [Ref. 14: p. 7].

Modeling has unique power and advantages. Analytical modeling has proven to be very useful in analyzing online, transaction-oriented systems difficult to analyze with simulation or other analysis methods. This may be economically effective for SPLICE use. It is particularly useful for estimating where bottlenecks will occur in a configuration. It provides an overall structure to guide a CPE group logically from one problem area to another, and it provides a deeper understanding of an entire system. Furthermore, as opposed to being an instrument to assist in problem solving like the other tools, modeling is a way of directly solving a problem by allowing a total system or part of a system to be examined before making a major commitment to a system acquisition or modification. [Ref. 14: pp. 7-8]

The real advantages of analytic models are that they can generally be created in a short time, applied quickly, have no programming language limitations, consume relatively little computer time, are easily understood, and are essential when it is too expensive, too time-consuming, or too dangerous to experiment on the real system. There is no overhead issue since a model does not require an existing

system. This measurement tool does not get in the way of the task.

Despite the positive features of analytic modeling, there are some limitations. First of all, modeling may not be practical for studying a system which is not deterministic and, hence, validation of the model against actual measurements may be impossible. Secondly, when too many changes must be made to the independent variables in order to validate the model, perhaps the system is too complex for modeling. Thirdly, whenever elaborate models are created, a thorough knowledge of queueing theory is generally required unless a good commercial package that handles this can be obtained. Lastly, specific network modeling tools are not yet generally available [Ref. 18: p. 81]. Do not confuse these with network systems analyzers. Despite the limitations, however, IBM's Systems Management Institute stresses use of analytic queueing models in computer-oriented performance evaluation classes [Ref. 41: p. 325].

Chris Bailey writing for Electronic Design magazine asserts:

> "The best modeling approaches are based on a combination of analytical and simulation techniques." [Ref. 26: p. 206]

It might be of some interest that research in the use of petri nets for modeling systems which have events occurring concurrently but with constraints on the concurrence, precedence, or frequency of the occurrences and in the performance evaluation of distributed systems are available. Use of this modeling technique has shown some utility in discovering overloads on a system, peak workloads, and bottlenecks. [Ref. 66,: p. 223, 83]67

LIST OF REFERENCES

1. Naval Supply Systems Command, Stock Point Logistics Integrated Communications Evironment (SPLICE) System Decision Paper III Executive Summary, by SPLICE Project Office (SUP 0472), March 1, 1985.

2. Navy Fleet Material Support Office, FMSO Document No. F94L0-001-9260 FD-S001B, SPLICE Functional Description, March 14, 1983.

3. Automatic Data Processing Selection Office, Contract No. N66032-84-D-0002, SPLICE Contract to Federal Data Corporation, November 17, 1983.

4. Automatic Data Processing Selection Office, Contract No. N66032-82-0007, SPLICE Solicitation Document, ADPSO Project 80-80, March 1, 1982.

5. Navy Fleet Material Support Office, FMSO Document No. F94L0-001-9260 SS-SU01C, SPLICE System Specification, January 16, 1984.

6. Naval Supply Systems Command (NAVSUP), SPLICE Strategic Planning Document, by NAVSUP (Code 0472) SPLICE Project Office, October 15, 1984.

7. Radford, K. J., Strategic Planning: An Analytical Approach, Reston Publishing Co., Inc., 1980.

8. Katzan, H. A Manager's Guide to LAN's, Carnegie Press, Inc., 1983.

9. Ferrari, D., Serazzi, G., and Zeigner, A., Measurement and Tuning of Computing Systems, Prentice-Hall, Inc., 1983.

10. Borovits, I. and Neumann, S., Computer Systems Performance Evaluation: Criteria, Measurements, Techniques, and Costs, Lexington Books, 1979.

11. Machlin, R. N., "Managing a Local Area Network", TELECOMMUNICATIONS, v. 18, no. 11, November 1984.

12. Fitzpatrick, M., "A Cure for Trial-and-Error Network Management", TELECOMMUNICATIONS, January 1985.

13. Stallings, W., LOCAL NETWORKS, An Introduction, Macmillan Publishing Company, 1984.

14. Morris, M. F. and Roth, P. F., Computer Performance Evaluation Tools and Techniques for Effectiveness Analysis, Von Nostrand Reinhold Company, 1982.

15. Abrams, M. D., "A New Approach to Performance Evaluation of Computer Networks", Computer Networking, edited by R. P. Blanc and I. W. Cotton for IEEE Press, 1976.

16. Holub, D., "Measuring Systems Performance", Auerbach Series in Data Processing Management, No. 5-03-03, March, 1980.

17. Amer, P. D. and Goel, A. K., "Performance Metrics for Bus and Token-Ring Local Area Networks", Journal of Telecommunications Networks, v. 2, no. 2, 1983.

18. Terplan, K., "Network Capacity Planning", Journal of Capacity Management, v. 2, no. 1, 1983.

19. Rosner, R. D., Packet Switching, Tomorrow's Communications Today, 371 pp., Lifetime Learning Publications, A Dvision of Wadsworth, Inc., 1982.

20. Fleet Material Support Office, Report No. NPS 54-82-003, Functional Design of a Local Area Network for the Stock Point Logistics Integrated Communications Environment, by N. F. Schneidewind of Naval Postgraduate School, December 1982.

21. Rajaraman, M. K., "Performance Measures for a Local Network", Performance Evaluation Review, v. 12, no. 2, Spring-Summer 1984.

22. Clark, D. D., Pogran, K. T., and Reed, D. P., "An Introduction to Local Area Networks", Tutorial on Local Computer Networks, COMPCON Fall '80, edited by K. J. Thurber and H. A. Freeman for Institute of Electrical and Electronics Engineers (IEEE), November 1978.

23. Frank, H., "Broadband versus Baseband Local Area Networks", TELECOMMUNICATIONS, v. 17, no. 3, March 1983.

24. Way, D. E., "Managing a LAN", TELECOMMUNICATIONS, v. 18, no. 1, January 1984.

25. Kee, K. C. E., Intro to Local Area Computer Networks, John Wiley & Sons, Inc., 1983.

26. Bailey, Chris, "Special Series on System Integration", Electronic Design, May 12, 1983.

27. Watson, W. B., "Configuration-Dependent Performance of a Prioritized CSMA Broadcast Network", COMPUTER, v. 14, no. 2, February 1981.

28. Franta, W. R. and Chlamtac, I., Local Networks, Lexington Books, D. C. Heath and Co., 1981.

29. Abrams, M. D., "Observations on Operating a Local Area Network", COMPUTER, May 1985.

30. Dhas, C. R. and Konangi, V. K., "Performance Parameters for a Packet Switched Network", IEEE 1985 Phoenix Conference on Computers and Communications, March 20-22, 1985.

31. Abrams, M. D. and Treu, S., "A Methodology for Interactive Computer Service Measurement", Communications of the ACM, v. 20, no. 12, December 1977.

32. National Bureau of Standards Technical Note 882, Criteria for the Performance Evaluation of Data Communications Services for Computer Networks, by D. S. Grubb and I. W. Cotton, September 1975.

33. Wiggins, R., "Intelligent Networking", TELECOMMUNICATIONS, v. 17, no. 1, January 1983.

34. Keifer, M., "Network Management", TELECOMMUNICATIONS, v. 18, no. 1, January 1984.

35. Grubb, D. S. and Cotton, I. W., "Rating Performance", Computer Networks: A Tutorial, 3rd. edition, edited by M. D. Abrams, R. P. Blanc, and I. W. Cotton, 1980.

36. Sussenguth, E. H., "Progress in Computer Networks", Information Processing '83, edited by R. E. A. Mason for Proceedings of the IFIP 9th World Computer Congress (Paris, France), September 1983.

37. Johnson, J. T., "Universal Flow and Capacity Index Gives Picture of Network Efficiency", DATA COMMUNICATIONS, February 1985.

38. Schneidewind, N. F., "Internetconnecting Local Networks to Long-Distance Networks", COMPUTER, September 1983.

39. Cortada, J. W., Managing DP Hardware: Capacity Planning, Cost Justification, Availability, and Energy Measurement, John Wiley & Sons, Inc. 1983.

40. Hopewell, L., "Management Planning in the Data Communications Environment", AFIPS Conference Proceedings of National Computer Conference, v. 43, 1974.

41. Allen, A. O., "Capacity Planning for Management", Proceedings of the 1983 Computer Measuremetn Group International Conference, December 6-9, 1983.

42. Leach, J. R., "Installation and Management of a Modern Communications Network", Proceedings of the 1982 Computer Measurement Group International Conference, 1982.

43. Buzin, J. P., "Use of Models for Capacity Planning", Computer Performance Evaluation Users Group (CPEUG) 14th. Meeting, edited by James Weatherbee for U. S. Department of Commerce, National Bureau of Standards, October 1978.

44. Mohr, J. M., "Projecting Workloads for New Systems: A Management Introduction", Journal of Capacity Management, v. 2, no. 1984.

45. Tandem Computers Incorporated, Tandem Nonstop (TM) Systems XRAY User's Manual, December 1983.

46. Tandem Computers Incorporated, Part No. 82003 C00, Introduction to Tandem Computer System, December 1983.

47. Datapro Research Corporation, Report No. M11-822-101, Tandem Nonstop Systems, September 1983.

48. Tandem Computers Incorporated, Report No. 109001-0983, Tandem Nonstop TXP System, Hardware Architecture, 1983.

49. Tandem Computers Incorpcrated, Part No. 82373 A00, Introduction to the Tandem 6100 Communications Subsystem, December 1983.

50. Tandem Computers Incorpcrated, Part No. 82311 B00, Introduction to Tandem Data Communications, December 1983.

51. Abdou, E., "Performance Analysis of Front-End and Host Processor Interface Configurations", Performance of Computer Installations, edited by D. Ferrari, 1978.

52. Carson, J. H. and Forman, E. H., "Analysis of Local Area Network Performance", IEEE 1981 Computer Networking, 1981.

53. Thornton, J. E., "Overview of HYPERchannel", COMPCON SPRING '79 Digest of Papers, March 1979.

54. Franta, W. R. and Heath, J. R., "Measurement and Analysis of HYPERchannel Networks", IEEE TRANSACTIONS CN COMPUTERS, v. C-33, no. 3, March 1984.

55. Thornton, J. E. and Christensen, G. S., "HYPERchannel Network Links", COMPUTER, September 1985.

56. Conservation with Mr. Mars Muller of Naval Supply Center (NSC), Oakland, California on July 1, 1985.

57. Telephone conversation with Mrs. Dottie Rogers of Naval Supply Center (NSC), Oakland, California on September 1985.

58. Telephone conversation with Mr. Bruce Alchorn of Naval Supply Center (NSC), Oakland, California on September 9, 1985.

59. Aupperle, E. M., "Merit's Evolution--Statistically Speaking", IEEE TRANSACTIONS ON COMPUTERS, v. C-32, no. 10, October 1983.

60. Pawlita, P. F., "Traffic Measurements in Data Networks, Recent Measurements, Results, and Some Implications", IEEE TRANSACTIONS ON COMMUNICATIONS, v. COM-29, no. 4, April 1981.

61. Defense Communications Agency, Defense Data Network

62. Naval Telecommunications Command Letter 2070: Ser. N5/7639 to Chief of Naval Operations, Subject: Defense Data Network (DDN) Waiver Extension, July 5, 1985.

63. VonTaube, E., "Internetworking: Connecting LAN's", TELECOMMUNICATIONS, v. 18, no. 10, 1984.

64. Defense Communications Agency, Defense Data Network

65. Opel, C. E., Network Management of the SPLICE Computer Network, MS Thesis, Naval Postgraduate School, Monterey, California, December 1982.

66. Peterson, J. L., "Petri Nets", Computing Surveys, v. 9, no. 3, September 1977.

67. Han, Y. W., "Performance Evaluation with Petri Nets", Computer Performance Evaluation Users Group (CPEUG) 14th. Meeting, edited by James Weatherbee for U.S. Department of Commerce, National Bureau of Standards, October 1978.

INITIAL DISTRIBUTION LIST

No.		Copies
1.	Defense Technical Information Center Cameron Station Alexandria, Virginia 22304-6145	2
2.	Superintendent Attn: Library Code 0142 Naval Postgraduate School Monterey, California 93943-5100	2
3.	Computer Technology Programs, Code 37 Naval Postgraduate School Monterey, California 93943-5100	1
4.	Professor Norman F. Schneidewind Code 54Ss Administrative Sciences Department Naval Postgraduate School Monterey, California 93943-5100	1
5.	LCDR Barry A. Frew Code 54FW Administrative Sciences Department Naval Postgraduate School Monterey, California 93943-5100	1
6.	Commander Naval Supply Systems Command Attn: CDR Dana Fuller, SC, USN (SUP 043) Washington, DC 20376	1
7.	Commanding Officer Navy Fleet Material Support Office Attn: LCDR Ron Nichols, SC, USN (Code 9RL) 5450 Carlisle Pike P. O. Box 2010 Mechanicsburg, Pennsylvania 17055-0787	1
8.	Director Defense Communications Agency 8th & South Courthouse Roads Washington, DC 20305	1
9.	Commanding Officer Navy Ships Parts Control Center P. O. Box 2020 Mechanicsburg, Pennsylvania 17055	1
10.	Commanding Officer Naval Supply Center Oakland, California 94625	1
11.	LCDR David D. Blankenship, USN Computer Technology Programs, Code 371 Naval Postgraduate School Monterey, California 93943-5100	1

12. ICDR Jonathan B. Schmidt, USN 1
Executive Officer
USS O'CALLAHAN (FF 1051)
FPO San Francisco, California 96674

13. ICDR Stephen M. Carr, SC, USN 1
Navy Management Systems Support Office
Naval Air Station
Norfolk, Virginia 23511-6694

14. Commander 1
Naval Supply Systems Command
Attn: CDR Arden Goss, SC, USN (SUP 0472)
Washington, DC 20376

15. Commander 1
Naval Supply Systems Command
Attn: Ms. Linda Matthews (SUP 0451)
Washington, DC 20376

16. ICDR Ted Case, SC, USN 1
SMC Box 1153
Naval Postgraduate School
Monterey, California 93943

END

FILMED

12-85

DTIC